NASA TECHNICAL NOTE

NASA TN D-6388
$C.1$

NASA TN D-6388

EQUIVALENT MEASUREMENTS,
OBSERVABILITY, AND CALIBRATION
ACCURACY LIMITS IN LINEAR ESTIMATION

by Luigi S. Cicolani

Ames Research Center
Moffett Field, Calif. 94035

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MAY 1971

1. Report No. NASA TN D-6388	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle EQUIVALENT MEASUREMENTS, OBSERVABILITY, AND CALIBRATION ACCURACY LIMITS IN LINEAR ESTIMATION		5. Report Date May 1971
		6. Performing Organization Code
7. Author(s) Luigi S. Cicolani		8. Performing Organization Report No. A-3442
		10. Work Unit No. 125-17-05-08-00-21
9. Performing Organization Name and Address NASA Ames Research Center Moffett Field, Calif., 94035		11. Contract or Grant No.
		13. Type of Report and Period Covered Technical Note
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D. C., 20546		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The structure of the linear estimation problem is investigated. First, data compression, performance analysis, and information separation are examined in the case of processing arbitrary sets of measurements. Second, the measurements are limited to samplings of given data types so that properties of the data types are common to any sampling. Thus, the unknowns to be estimated can be reduced to those observable to the data types. Finally, estimation is intended to compare an unknown system to internationally defined physical units. Measurements separate into comparisons with physical units (calibrations) and comparisons among unknowns. Generally, measurements are of the second type and their processing compares the unknown system to intermediate physical standards which appear as unobservable parameters that limit the achievable estimation accuracy.

17. Key Words (Suggested by Author(s)) Linear estimation Observability Kalman filter Calibration accuracy Measurement compression	18. Distribution Statement Unclassified — Unlimited		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 83	22. Price* $3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

Most symbols are defined and used ad hoc in the text. Script capitals are linear vector spaces, capitals are matrices, except where otherwise indicated, and vectors are indicated by the superscript $\frown$ or defined as such. Other symbols appearing routinely are listed below.

$(\ )^{-1}, (\ )^{\#}, (\ )^{T}$	inverse, pseudo inverse, and transpose of $(\ )$
$(\hat{\ }), (\tilde{\ })$	estimate and error in estimate of $(\ )$
$\doteq$	equivalence symbol
$\cap$	intersection of two sets
$\oplus$	direct sum of two linear vector spaces
$\mathcal{L}\{\ \}$	linear vector space spanned by contents of $\{\ \}$ Where a matrix appears, the column space of the matrix is indicated.
h	vector (costate) defining a parameter as a linear combination of the state variables
$\mathcal{H}$	parameter (costate) space
$\mathcal{H}_m, \mathcal{H}_u$	observable and unobservable subspaces of $\mathcal{H}$
H	matrix whose columns are the costates of a set of measurements
H_m, H_u	matrices whose columns are bases of $\mathcal{H}_m, \mathcal{H}_u$
MS	mean square value
m(X)	data type, given as a scalar function of the state
P, P_o, P_A	posterior, prior, and initial covariances of the state estimation error
q	MS measurement noise for a measurement
Q	diagonal matrix of MS measurement noise for a set of measurements
(H,Q)	measurement sequence (the costates and noise matrix of a set of measurements)
t	time
X(t)	state vector

$\{X_A(t), t_o \leqslant t \leqslant t_F\}$ reference solution of the state equation

$x(t)$ state deviation vector, $X(t) - X_A(t)$

χ state space

χ_m, χ_u observable and unobservable subspaces of χ

X_m, X_u matrices whose columns are bases of χ_m, χ_u

w transformed state deviation vector

W, W_o, W_A posterior, prior, and initial covariances of $\tilde{w}$

y data vector for a set of measurements

$\Phi(t_2, t_1)$ transition matrix of the linear state equation

EQUIVALENT MEASUREMENTS, OBSERVABILITY, AND CALIBRATION

ACCURACY LIMITS IN LINEAR ESTIMATION

Luigi S. Cicolani

Ames Research Center

SUMMARY

Some aspects of the structure of the linear estimation problem are investigated in this report. In the first section, the processing of arbitrary sets of measurements is considered. The notion of equivalent sets of measurements is defined and applied to obtain a unified view of such operations as data compression, performance analysis, and selective processing of data.

In the second section, the measurements to be processsed are assumed to be samplings, obtained at some arbitrary sequence of times, of one or more given data types or time-varying functions of the unknowns. The properties of the data types are then common to any sampling. One such property, observability, is well known. The system is observable to the data types, provided all unknowns can be determined from some sampling of the data types. Otherwise, only some lesser number of linear combinations of the unknowns can be determined and the estimation problem can be reduced to one of estimating this smaller set of unknowns.

Finally, the physical nature of measurements and estimation is considered in combination with observability analysis to obtain a theory for the calibration accuracy limits inherent in estimation problems. All measurements are comparisons of like physical phenomena and the basis of comparison is made unique by international definition of physical units. The estimation process combines measurements in order to compare some physical system, as defined by a set of unknown state variables, with these physical units. Measurements separate into comparisons with physical units (calibrations) and comparisons among unknowns (hence, are nonlinear in the unknowns). If all calibrations are carried out in an initial set of measurements, as is usually the case, then the processing of subsequent measurements becomes a comparison of the system with a set of intermediate physical standards. These standards appear as unobservable parameters in the problem and the accuracy to which they are estimated from the initial set of measurements is a lower bound, or calibration accuracy limit, on the accuracy to which the system can be estimated without performing new calibrations. Intermediate standards for quantities having composite units, such as velocities, can also be defined from the elementary standards, and these are applied to examine the estimation of the speed of light as part of space vehicle orbit estimation problems.

INTRODUCTION

Linear estimation is the problem of computing the best estimate of a set
of unknowns from noisy measurements of linear functions of the unknowns. This
subject has received wide attention in the engineering literature in recent
years and has applications in many fields. A particularly complex example is
the estimation of space vehicle orbits where accuracy requirements have led
to the development of radar station networks capable of generating vast
amounts of data and of sophisticated data processing programs which estimate
many unknowns. In these and other large-order estimation problems, physical
insight is difficult to achieve and there is little theory of a general or
routine nature for analyzing the relations among the basic factors in the
problem and the estimation performance, or for analyzing the causes of such
common difficulties as ill-conditioning of the information provided by the
measurements. Such difficulties are often characteristic of the measurements
themselves, but with the exception of observability (ref. 1), few measurement
characteristics of a general nature have been identified and observability has
itself received limited attention in practice.

This report is concerned with the analysis of measurements and the
relation of measurement characteristics to the resulting estimation perfor-
mance. Past work on such relations has been largely empirical or restricted
to particular problems. For example, in orbit estimation problems one can
simulate the estimation problem and then calculate the effect of various types
of measurements, measurement schedules, orbit geometry, etc., on performance
(e.g., refs. 2, 3). In several problems, analysis of the equations relating
the data to properly selected orbit parameters has clarified some aspects of
performance and computational procedure for these cases (refs. 4, 5, 6).

The analysis is carried out in the context of the linear estimation
problem, including assumptions of unforced dynamics and uncorrelated gaussian
measurement noise. The first section deals with arbitrary sequences of
measurements. It is found that a group of "equivalent sequences" can be
derived from the original set of measurements, each of which gives the same
state estimate and performance. Equivalent sequences permit a number of
operations on the data and the processing equations; the measurements can be
compressed to an equivalent sequence of minimum size, certain equivalent
sequences give the performance, and others permit selective processing of
information that can be used to separate the estimation calculations into
independent lower order parts by selectively deleting some information.

The next section deals with characteristics of the data type.
Measurement sequences are not always arbitrary sets of measurements but are
often obtained by sampling one or more data types at some arbitrary sequence
of times. The data type, given as a function of the state, is then a con-
straint on the measurements, and its characteristics influence estimation
performance independent of the particular sample taken and processed. One
such characteristic is observability; that is, the ability to determine the
state from some (non-noisy) sampling of the data type (refs. 1, 7, 8). If the
state is indeterminate from the data type, then unobservable states can be
defined but a part of the state cannot be estimated from any sampling of the

data type. In that case, the order of the data processing equations can be reduced by the number of independent unobservable states, and a lower bound on the estimation accuracy obtainable from the data type is given by the initial estimation errors associated with the unobservable states.

In the third section, a theory of calibration accuracy in estimation problems is obtained as an application of observability analysis. Estimation is an exercise in comparing a physical system to internationally defined unit quantities (meter, kilogram, etc.). In most problems direct comparisons with these unit quantities are carried out in an initial set of measurements and the data types are comparisons among functions of the unknowns. The estimation procedure can then be shown to be one of comparing the physical system to intermediate standards of length, mass, etc. These standards are unobservable parameters in the problem and their initial estimation accuracies provide lower bounds on the accuracy achievable by processing any sampling of the data types. An intermediate standard for velocities derived from the basic standards is applied to examine estimation of the speed of light as part of the orbit estimation problem.

BASIC FORMULAS FOR THE LINEAR ESTIMATION PROBLEM

The well-known state estimation procedure reviewed below is based on the assumptions that (1) the state is governed by unforced dynamics which are either linear or can be described by equations linearized about a reference state, and (2) finite sequences of measurements are made, subject to uncorrelated gaussian-distributed noise with zero mean, and processed to obtain the minimum variance estimate of the state.

The state (an $n \times 1$ vector) of a linear system satisfies an equation of the form

$$\dot{x}(t) = F(t)x(t) \tag{1}$$

Here, $F(t)$ is a known continuous $n \times n$ matrix. These are unforced dynamics so that the time-history of the state is uniquely defined by the initial conditions $x(t_o)$, as is expressed in the usual solution of (1)

$$x(t) = \Phi(t,t_o)x(t_o) \tag{2}$$

where $\Phi(t,t_o)$ is the transition matrix of the system.

A scalar output of the system is available for sampling during some time period, $[t_o,t_F]$, and is assumed given in the form

$$m(t) = \bar{\mu}(t)^T x(t) \qquad t_o \leqslant t \leqslant t_F \tag{3a}$$

or, alternatively

$$m(t) = h^T(t)x(t_o) \qquad t_o \leqslant t \leqslant t_F \tag{3b}$$

where

$$h(t) = \Phi^T(t,t_o)\bar{\mu}(t)$$

3

The vector function $\{h(t), \; t_O \leqslant t \leqslant t_F\}$, termed the *measurement costate* hereafter, is known and continuous, and it defines the output as a linear combination of the components of the initial state. The state dynamics are absorbed in the definition of $h(t)$. If more than one scalar output is available, then additional relations of the form (3) are given.

The data to be processed in order to estimate the state are a (noisy) sampling of the output at some finite sequence of times, $t_1, \ldots, t_N$, in $[t_O, t_F]$. These data are related to the state by

$$y_i = h^T(t_i)x(t_o) + \varepsilon_i \qquad i = 1, \ldots, N \qquad (4)$$

where $\{y_i\}$ are the actual data and ε_i is the measurement noise, assumed to be a gaussian distributed random variable with zero mean and known nonzero MS value, $q(t_i)$. The sample times are assumed sufficiently separated for the noise to be uncorrelated among the measurements. The notation for the costates, noise variances, and data from any set of measurements is abbreviated to

$$H = [h(t_i) \; . \; . \; . \; h(t_n)] \qquad Q = \begin{bmatrix} q(t_i) \end{bmatrix} \qquad y = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_N \end{pmatrix}$$

It will also be convenient to refer to the two matrices, (H, Q), as a *measurement sequence*. In the first section of this report, the costates of a measurement sequence, H, are taken as arbitrary rather than constrained to samplings of a given output function of the form (3). The second section considers the effects of introducing such a constraint.

This report deals with the processing of measurements with the sequential minimum variance equation (ref. 9). Prior measurements are assumed to have been processed and to have yielded the estimate, $\hat{x}^-$, and covariance matrix

$$P_o = E[\tilde{x}^-(\tilde{x}^-)^T]$$

where $\tilde{x}$ is the estimation error $x - \hat{x}$. The new estimate and covariance obtained from processing (H, Q, y) are then

$$\left.\begin{aligned} \hat{x} &= P_oH[Q + H^TP_oH]^{-1}(y - H^T\hat{x}^-) + \hat{x}^- \\[2em] P &= [P_o^{-1} + HQ^{-1}H^T]^{-1} = P_o - P_oH[Q + H^TP_oH]^{-1}H^TP_o \end{aligned}\right\} \qquad (5)$$

The source of the prior estimate and covariance is unspecified here but it is assumed that P_o is positive definite.

In many problems as, for example, the orbit determination problem, the linear system, (1) - (4), is obtained as an approximation of a nonlinear one. Since consideration is given to the calibration accuracy limits in such linearized systems in a later section, a brief review of the linearization procedure is given here.

A nonlinear state equation and scalar output are assumed given in the form

$$\dot{X}(t) = \bar{f}(X(t)) \tag{6a}$$

$$m = m(X) \tag{6b}$$

The output, $m(X)$, and the components of $\bar{f}(X)$ are assumed defined on the state space and to have continuous partial derivatives. If more than one output is available for sampling, additional functions of the form (6b) are given. The state, X, is an $n \times 1$ vector list of independent quantities which suffice to define the system. Independence of the state variables means there is no nontrivial function, $g(X(t))$, that is zero at all times.

To linearize equations (6), first obtain a reference solution of (6a), denoted by

$$\{X_A(t), \quad t_0 \leqslant t \leqslant t_F\} \tag{7}$$

This solution is assumed determined as an unbiased estimate of $X(t_0)$ from some set of initial measurements which are arbitrary (not necessarily samplings of (6b)) and sufficient (they can be inverted to solve for $X(t_0)$). Define the error in this initial estimate

$$\tilde{x}(t_0) = X(t_0) - X_A(t_0)$$

where $X(t_0)$ is the unknown true state at t_0; then

$$E[\tilde{x}(t_0)] = 0 \tag{7a}$$

and the initial covariance

$$P_A = E[\tilde{x}(t_0)\tilde{x}^T(t_0)] \tag{7b}$$

is a (known) positive definite matrix.

To complete the linearization, define the state deviation

$$x(t) = X(t) - X_A(t) \qquad t_0 \leqslant t \leqslant t_F$$

The state deviation is assumed sufficiently "small" so that it satisfies the linearized state equation (1), with F(t) given by the gradient of (6a) with respect to the state variables

$$F(t) = \left[\frac{\partial f_i}{\partial X_j} \; i,j = 1, \; . \; . \; ., \; n \right]_{X_A(t)} \tag{8}$$

Here, the subscripts, i, j, denote the components of the vectors $\bar{f}(X)$ and X, respectively. Further, the linearized output (eq. (3)) is obtained from the gradient of equation (6b) by using

$$\bar{\mu}(t) = [\nabla m(X)]_{X_A(t)} \tag{9}$$

Equation (3) then gives the first-order difference between the output for the true state and the output for the reference state.

Finally, the estimation procedure is carried out by processing samplings of the output with equation (5). The processing is begun with the initial state estimate and covariance given by equations (7a) and (7b). These samplings can be processed in equation (5) in any convenient order - sequentially, in groups, or all at once. If the processing is done sequentially, then $\hat{x}^-$ and P_0 in equation (5) refer to the results of processing measurements prior to the current one.

EQUIVALENT SEQUENCES AND DATA

This section deals with the analysis of measurement sequences and estimation performance, that is, of equation (5). The restriction of measurement sequences to samplings of one or more data types is not imposed so that the discussion applies to arbitrary finite sequences. The principal tool in the analysis is the equivalent sequence and data, that is, any set of measurements and data (possibly fictitious) which, when processed, give the same state estimate and covariance as the real sequence and data. The equivalent sequence and data is used in the analysis to compress large sequences into small equivalent sequences, and to obtain equivalent sequences with special properties such as those which state the estimation performance and those which extract information on any specified set of parameters.

The basic quantities used in this section are defined next. A *parameter* is any linear combination of the state variables, $h^T x$, and is specified by a vector (costate), h, whose components are the constants of combination. If (H, Q, y) is any finite measurement sequence and the corresponding data vector, then the columns of H specify the parameters measured by the sequence and the column space of H is called the *observable parameter space* expressed as

$$\mathcal{H}_m = \mathcal{L}\{H\}$$

6

The *information matrix* for a sequence is

$$I_f \equiv HQ^{-1}H^T$$

A fictitious sequence, (H_E, Q_E), is said to be *equivalent* to the real sequence, (H, Q), if they have identical information matrices. That is, if

$$H_E Q_E^{-1} H_E^T = HQ^{-1}H^T$$

The noise matrix, Q_E, is assumed diagonal here. Both sequences give the same covariance when processed in equation (5).

If (H_E, Q_E) is equivalent to the real sequence (H, Q) and y is the real data, then the corresponding *equivalent data*, y_E, is that fictitious data vector such that (H_E, Q_E, y_E) yields the same estimate of the state in equation (5) as does (H, Q, y).

Equivalent Sequences

Any real sequence can be replaced by a variety of equivalent sequences, all of which have identical information. For example, the measurement noise can always be absorbed in the h vectors by defining

$$H' = HQ^{-1/2}$$

Then (H', I) is equivalent to (H, Q). Such sequences with unit mean square noise are called primary. It is apparent that any square root[1] of the information matrix is an equivalent primary sequence.

Sequences equivalent to (H, Q) can consist of r measurements, where r can be greater or less than the number of measurements in (H, Q). There is no upper limit, but a lower limit $r*$ exists and is

$$r* = \text{Rank}(H) \tag{10}$$

An equivalent sequence containing $r*$ measurements is called an *equivalent basis* since its costates are a basis of the column space of H.

A series of statements concerning the algebra of equivalent sequences follows next. Some minor proofs are omitted. The notation $(H_A, Q_A) \doteq (H_B, Q_B)$ means the two sequences are equivalent. Where primary sequences are involved, the noise matrix is dropped from this notation.

---

[1]The square root of a positive semidefinite matrix, M, refers to any matrix M_1 such that $M_1 M_1^T = M$.

[1] Equivalent sequences span the same space. If $H_A \doteq H_B$ then $\mathcal{L}\{H_A\} = \mathcal{L}\{H_B\}$.

[2] Let $[H_A \mid H_B]$ be a partitioned primary sequence. If $H_A \doteq H_C$ then $[H_A \mid H_B] \doteq [H_C \mid H_B]$.

[3] Suppose $[H_A \mid H_B] \doteq [H_C \mid H_D]$. Then $H_A \doteq H_C$ if and only if $H_B \doteq H_D$.

These statements are also true for nonprimary sequences.

[4] Every sequence has an equivalent basis.

Note that the information matrix is positive semidefinite. Any square root of I_f which has maximal rank is an equivalent *primary basis*. The mathematical literature may be consulted for various methods of computing square roots. One interesting method is described in reference 10 and is reviewed in appendix A.

[5] Let H_A be any equivalent primary basis of (H, Q). Then H_B is also an equivalent primary basis if and only if there exists an orthogonal $r^* \times r^*$ matrix, B, such that

$$H_A = H_B B$$

To prove [5], assume, first, that H_A is an equivalent primary basis of (H, Q) and is related to H_B by $H_A = H_B B$ where B is orthogonal. Then H_A, H_B are equivalent since they yield the same information matrix, whence H_B is also equivalent to (H, Q). Conversely, assume H_A, H_B, are equivalent primary bases of (H, Q). Then both are bases of the same subspace, $\mathcal{H}_m$, and are related by some nonsingular matrix, B, of size r^*; $H_A = H_B B$. Since H_A and H_B are equivalent, it follows that BB^T is the unit matrix, whence B is orthogonal.

[6] Let H_A, H_A' be any pair of primary equivalent bases of some sequence (H, Q), and let H_B, H_B' be any subsequences of H_A, H_A', respectively. Then $H_B \doteq H_B'$ if and only if $\mathcal{L}\{H_B\} = \mathcal{L}\{H_B'\}$.

To prove [6], note first, that if $H_B \doteq H_B'$ then they span the same space, (statement [1]). Conversely, suppose $\mathcal{L}\{H_B\} = \mathcal{L}\{H_B'\}$. Since H_A, H_A' are equivalent primary bases, then (statement [5]) they are related by an orthogonal matrix, B.

$$H_A' = H_A B$$

This equation can be partitioned as

$$\left[H_B' \mid H_R' \right] = \left[H_B \mid H_R \right] \begin{bmatrix} B_1 & \mid & B_2 \\ -- & + & -- \\ B_3 & \mid & B_4 \end{bmatrix}$$

8

or

$$H_B' = H_B B_1 + H_R B_3$$

Since H_B, H_B' are both bases of the same space then B_3 is zero. Since B was orthogonal then, B_1 is orthogonal and H_B, H_B' are equivalent (statement [5]).

As seen in [5], many equivalent basis sequences are available and several with special properties are discussed in later sections. Basis sequences compress the original set of measurements to $r*$ measurements which can be used in the data processing in place of the real sequence.

Equivalent Data

Suppose (H_E, Q_E) is any sequence equivalent to the real sequence (H, Q). The corresponding equivalent data vector, y_E, is that fictitious data vector for which the state estimate, $\hat{x}$, obtained from processing (H_E, Q_E, y_E) in equation (5), is identical to the estimate obtained from (H, Q, y).

The equivalent data are more easily derived from an alternate formula for the state estimate. Let

$$H_T = [H_0 \vdots H], \qquad Q_T = \begin{bmatrix} Q_0 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & Q \end{bmatrix}, \qquad y_T = \begin{pmatrix} y_0 \\ \cdots \\ y \end{pmatrix}$$

that is, (H_T, Q_T, y_T) is the total set of all prior and current measurements. The minimum variance estimate of the state from the total set of measurements is (ref. 9)

$$\hat{x} = [H_T Q_T^{-1} H_T^T]^{-1} H_T Q_T^{-1} y_T$$

This estimate is identical to that given by equation (5). The vector y_E is defined by the condition that the state estimate be unchanged when (H, Q, y) is replaced by (H_E, Q_E, y_E) in the above formula. Since $H_T Q_T^{-1} H_T^T$ is the total information matrix, which is nonsingular and which is unchanged if any subsequence is replaced by an equivalent subsequence (statement [2]), then y_E satisfies

$$[H_0 \vdots H_E] \begin{bmatrix} Q_0 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & Q_E \end{bmatrix}^{-1} \begin{pmatrix} y_0 \\ \cdots \\ y_E \end{pmatrix} = [H_0 \vdots H] \begin{bmatrix} Q_0 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & Q \end{bmatrix}^{-1} \begin{pmatrix} y_0 \\ \cdots \\ y \end{pmatrix}$$

or, after expansion,

$$H_E Q_E^{-1} y_E = H Q^{-1} y$$

If (H_E, Q_E) is an equivalent basis, then there is a unique solution for y_E since every column of H can then be given as a unique linear combination of the columns of H_E; that is, H can be given as

$$H = H_E M \qquad (11)$$

Substitute this in the preceding equation and premultiply both sides by $H_E^T P_o$ to obtain

$$y_E = Q_E M Q^{-1} y \qquad (12a)$$

valid for equivalent basis sequences. The result depends only on the real sequence and data and is independent of prior measurements. Alternatively, the numerical pseudo-inverse (ref. 11) of H_E is

$$H_E^{\#} = (H_E^T H_E)^{-1} H_E^T$$

whence the solution for y_E can be given as

$$y_E = Q_E H_E^{\#} H Q^{-1} y \qquad (12b)$$

valid for equivalent basis sequences. The operations $H_E^T H_E$, $H_E^T H$ that appear in this solution are often physically undefined owing to the physical units commonly associated with components of h vectors, but the result in equation (12b) is numerically correct since the physically undefined operations appear in $H_E^{\#} H$ in cancelling pairs. In fact, the product $H_E^{\#} H$ is the well-defined matrix M of equation (11).

The Uncorrelated Equivalent Basis Sequence

An *uncorrelated equivalent basis* of the real sequence (H, Q) is any primary equivalent basis H_E for which $H_E^T P_o H_E$ is diagonal. The costates of such a sequence define a set of parameters

$$\{h_j^T x, \qquad j = 1, \ldots, r^*\}$$

whose estimation errors are uncorrelated both before and after (H, Q) is processed. It is shown that every real sequence has such a basis and it is usually unique, given the real sequence and prior covariance. Further, the elements of the diagonal matrix state the performance in reducing the mean square estimation errors from their prior values specified by P_o.

10

[7] The uncorrelated equivalent basis. Every sequence has an uncorrelated equivalent basis given as follows: Let H_E be any equivalent basis of (H, Q) and let P_O be the prior covariance matrix. Define the matrix

$$A = H_E^T P_O H_E$$

and let Λ, B be its eigenvalues and modal matrix. Then the matrix

$$H_{EU} = H_E B$$

is an uncorrelated equivalent basis of (H, Q) and

$$H_{EU}^T P_O H_{EU} = \Lambda$$

$$H_{EU}^T P H_{EU} = \Lambda [I + \Lambda]^{-1}$$

To prove [7], note that P_O is positive definite and symmetric, and that H_E always exists and is an $n \times r^*$ matrix of rank r^*. Then A is a positive definite symmetric matrix of size r^*. Let its eigenvalues and modal matrix be Λ, B; that is, the columns of B are the eigenvectors of A taken in the same order as the eigenvalues in the diagonal matrix Λ. The matrix B is orthogonal due to the symmetry of A and it then follows from statement [5], that $H_E B$ (i.e., H_{EU}) and H_E are equivalent primary sequences. Therefore, $H_E B$ is equivalent to (H, Q), is a basis, and makes P_O diagonal. Of the final two equations in statement [7], the first is true by construction of H_{EU}, and the second follows readily from equation (5) after (H, Q) is replaced by its equivalent (H_{EU}, I), and $H_{EU}^T P H_{EU}$ is formed.

The r^* columns of H_{EU} are costates which define r^* parameters whose prior estimation errors are mutually uncorrelated and whose posterior errors are also mutually uncorrelated. This follows since, as a result of statement [7]

$$h_i^T P_O h_j = h_i^T P h_j = 0$$

for h_i, h_j any two distinct columns of H_{EU}. Further, after processing (H, Q) the MS estimation errors of the parameters defined by H_{EU} are reduced by the respective factors $\{(1 + \lambda_i)^{-1}, i = 1, \ldots, r^*\}$.

The uncorrelated basis is a unique characteristic of a sequence (H, Q), and the *prior covariance*, P_O.

[8] Uniqueness of eigenvalues. Given (H, Q) and P_O, the eigenvalues of the matrix

$$A(H_E) \equiv H_E^T P_O H_E$$

are *invariant for all primary equivalent bases, H_E, of (H, Q).*

11

Noting statement [5], all such matrices, $A(H_E)$, are related by orthogonal transformations and, therefore, have identical eigenvalues.

[9] Uniqueness of the uncorrelated equivalent basis. Given (H, Q) and P_o, the uncorrelated equivalent basis H_{EU} defined by [7] has as many unique columns, except for ordering and sign, as A has distinct eigenvalues.

Distinct eigenvalues are those which appear in Λ only once. The occurrence of multiple eigenvalues is accidental so that normally H_{EU} is unique except for ordering and sign. To prove [9], let H_{E1}, H_{E2} be distinct equivalent bases of (H, Q). These are related by some orthogonal matrix C (statement [5]) so that

$$A(H_{E1}) = C^T A(H_{E2}) C$$

If B_1 and B_2 are the modal matrices of $A(H_{E1})$ and $A(H_{E2})$, respectively, then

$$\Lambda = B_1^T C A(H_{E2}) C B_1 = B_2^T A(H_{E2}) B_2$$

Therefore, both B_2 and CB_1 are modal matrices of $A(H_{E2})$. If the eigenvalues Λ are distinct, the modal matrix of $A(H_{E2})$ is unique except for the signs of its columns. If signs are ignored,

$$B_2 = CB_1$$

then

$$H_{EU2} \equiv H_{E2} B_2 = H_{E2} C B_1 = H_{E1} B_1 \equiv H_{EU1}$$

that is, the uncorrelated bases obtained from H_{E1}, H_{E2} are identical. If Λ has an eigenvalue of multiplicity s, the corresponding s eigenvectors can be any orthogonal basis of the same s-dimensional eigenspace. In that case, s columns of H_{EU1} and H_{EU2} need not be identical, although they are equivalent, while the remaining columns are identical.

In the next section, the uncorrelated equivalent basis gives the performance with which (H, Q) reduces estimation errors.

Observability and Estimation Performance

Performance, in this section, refers to the reduction of MS estimation errors obtained from processing a sequence. The analysis is based on the decomposition of the state and costate (dual of χ) spaces into subspaces termed observable and unobservable to the sequence. Estimation errors and the covariance matrix are similarly decomposed, and the effect of processing the

sequence is to reduce only the observable errors, the amount of reduction being stated by the uncorrelated basis sequence and its eigenvalues.

Decomposition of $\mathcal{H}$ *and* X- The n-dimensional costate space $\mathcal{H}$ and state space X are partitioned into the following subspaces whose definitions are based on the costates, h, of the measurements to be processed and on the (nonsingular) prior covariance, P_o:

$$\mathcal{H}_m \equiv \mathcal{L}\{H\} \tag{13a}$$

$$\mathsf{X}_u \equiv \{x: \quad x^T H = 0\} \tag{13b}$$

$$\mathcal{H}_u \equiv \{h: \quad H^T P_o h = 0\} \tag{13c}$$

$$\mathsf{X}_m \equiv \mathcal{L}\{P_o H\} \tag{13d}$$

The subspace $\mathcal{H}_m$ is the r*-dimensional column space of H - the space spanned by the measurement costates - and is termed the observable parameter space. The subspace X_u is an (n - r*)-dimensional subspace of X, termed the unobservable states. These two subspaces are properties of only the measurement costates and are well known in discussions of observability (e.g., ref. 7).

The remaining two subspaces $\mathcal{H}_u$ and X_m are termed the unobservable parameters and observable states, respectively, and are defined from the prior covariance as well as the costates of the measurement sequence.

The subspaces defined in equations (13) provide a complete decomposition of $\mathcal{H}$ and X since they satisfy

$$\mathcal{H} = \mathcal{H}_m \oplus \mathcal{H}_u \qquad \mathcal{H}_m \cap \mathcal{H}_u = \{0\}$$

$$\mathsf{X} = \mathsf{X}_m \oplus \mathsf{X}_u \qquad \mathsf{X}_m \cap \mathsf{X}_u = \{0\}$$

In that case any costate h or state x can be uniquely decomposed into a sum

$$h = h_m + h_u$$

$$x = x_m + x_u$$

whose parts are, respectively, in $\mathcal{H}_m$, $\mathcal{H}_u$ or X_m, X_u. Further, if H_m, H_u are two matrices whose columns are bases of $\mathcal{H}_m$, $\mathcal{H}_u$, respectively, then related bases of X_m, X_u are given by the columns of the matrices

$$\mathsf{X}_m \equiv P_o H_m, \qquad \mathsf{X}_u \equiv P_o H_u \tag{14}$$

The decomposition of $\mathcal{H}$ and X defined in equations (13) is "orthogonal" in the sense that the inner products

$$\langle h_1, h_2 \rangle \equiv h_1^T P_o h_2$$

$$\langle x_1, x_2 \rangle \equiv x_1^T P_o^{-1} x_2$$

are zero for all pairs (h_1, h_2) or (x_1, x_2) having one member in $\mathcal{H}_m$ (or X_m) and the other in $\mathcal{H}_u$ (or X_u). For the inner product on $\mathcal{H}$, a "distance" $\sqrt{\langle h, h \rangle}$ is, physically, the RMS prior estimation error of the parameter defined by h. Two costates are orthogonal

$$\langle h_1, h_2 \rangle = 0$$

provided the prior estimation errors of the two parameters defined by h_1 and h_2 are uncorrelated. Thus, the estimation error for any parameter whose costate is in $\mathcal{H}_u$ is uncorrelated with the estimation error for any parameter whose costate is in $\mathcal{H}_m$, including the measured parameters, and, hence, cannot be reduced by processing the measurements.

The usual inner products, $h_1^T h_2$, $x_1^T x_2$ are often physically undefined in estimation problems because of the physical units associated with state variables.

It may be noted that the data are independent of that part of the state which is in X_u and depends only on that part which is in X_m. As noted earlier, the state can be uniquely decomposed

$$x = x_m + x_u$$

and since, by construction, all measurement costates h_m are in $\mathcal{H}_m$, we have, using equation (13b)

$$h_m^T x = h_m^T x_m$$

Reduction of estimation errors- The state estimation error, $\tilde{x}$, can be uniquely decomposed into parts in X_m and X_u

$$\tilde{x} = \tilde{x}_m + \tilde{x}_u$$

If, next, H_m, H_u are any bases of $\mathcal{H}_m$, $\mathcal{H}_u$, and X_m, X_u are defined in accordance with equation (14), then the prior covariance of $\tilde{x}$ can be given by

$$P_o = X_m [H_m^T P_o H_m]^{-1} X_m^T + X_u [H_u^T P_o H_u]^{-1} X_u^T \tag{15}$$

14

or, simply

$$P_o = P_m + P_u$$

This equation is an identity obtained by defining the matrices

$$\theta \equiv \left[H_m \mid H_u \right] \quad \text{and} \quad \Psi \equiv \left[X_m \mid X_u \right] \equiv P_o \theta$$

where θ and Ψ are $n \times n$ nonsingular matrices by construction. Next, form the identity

$$P_o = P_o \theta \theta^{-1} P_o^{-1} \theta^{-T} \theta^T P_o = \Psi \left[\theta^T P_o \theta \right]^{-1} \Psi^T$$

which, after expanding and noting equation (13c), yields equation (15).

Equation (15) gives P_o as a sum of two positive semidefinite matrices of rank r^* and $n - r^*$, respectively. This decomposition is unique in that the two matrices, P_m, P_u, are independent of the particular bases of $\mathcal{H}_m$ and $\mathcal{H}_u$ chosen for the construction. Further, it can be verified that these two matrices are, respectively, the prior covariances of $\tilde{x}_m$ and $\tilde{x}_u$.

If (H, Q) is some sequence to be processed and its uncorrelated equivalent basis is now chosen as the basis of $\mathcal{H}_m$ used in equation (15), the following is readily established.

[10] Let H_{Eu}, Λ be the uncorrelated basis and eigenvalues of (H, Q) and P_o, as defined in statement [7]. Let H_u be any basis of $\mathcal{H}_u$, and define

$$X_m \equiv P_o H_{EU}$$

$$P_u \equiv P_o H_u [H_u^T P_o H_u]^{-1} H_u^T P_o$$

Then the prior and posterior covariances are

$$P_o = X_m \Lambda^{-1} X_m^T + P_u$$

$$P = X_m \Lambda^{-1} [I + \Lambda]^{-1} X_m^T + P_u$$

The above expansion of P is obtained by using statement [7] and equation (15) in equation (5). As seen in statement [10], the covariance of $\tilde{x}_u$, given by P_u, is unreduced by processing (H, Q). The prior covariance of $\tilde{x}_m$ can be written

$$X_m \Lambda^{-1} X_m^T = \sum_{i=1}^{r^*} \frac{X_i X_i^T}{\lambda_i}$$

15

where $\{X_i\}$ are the columns of X_m here. The processing reduces each term in this sum by the respective factors $\{(1 + \lambda_i)^{-1}, i = 1, \ldots, r*\}$.

The *performance* in estimating any parameter, $h^T x$ is defined as

$$\rho = \frac{h^T P h}{h^T P_o h}$$

or the ratio of posterior and prior variances of $h^T \tilde{x}$. All values of ρ are in $(0, 1]$ and performance is better for smaller values.

A costate, h, can be given as the unique sum of its parts in $\mathcal{H}_m$ and $\mathcal{H}_u$

$$h = h_m + h_u = H_{EU} \bar{a} + h_u$$

where h_m has been given in terms of the basis, H_{EU}, of $\mathcal{H}_m$. Using statement [10], one obtains for any h

$$\rho = \frac{\displaystyle\sum_{i=1}^{r*} (\alpha_i/\lambda_i)(1 + \lambda_i)^{-1} + h_u^T P_u h_u}{\displaystyle\sum_{i=1}^{r*} (\alpha_i/\lambda_i) + h_u^T P_u h_u} \tag{16}$$

where $\{\alpha_i\}$ are the components of $\bar{a}$. From (16) it is apparent that

$$\frac{1}{1 + \lambda_{max}} \leqslant \rho \leqslant \frac{1}{1 + \lambda_{min}} \qquad \text{for} \quad h \varepsilon \mathcal{H}_m \tag{17}$$

$$\rho = 1 \qquad \text{for} \quad h \varepsilon \mathcal{H}_u$$

The extreme eigenvalues thus give the extremes of performance for parameters whose costates are in $\mathcal{H}_m$ while no reduction of MS estimation error occurs for parameters with costates in $\mathcal{H}_u$. More generally, if h is in neither subspace (h_m and h_u are both nontrivial), then performance is in the interval $(1/1 + \lambda_{max}, 1)$.

The uncorrelated basis is equivalent to the real sequence (H, Q) in the sense that the same estimate and performance would be achieved if the measurements defined by (H_{EU}, I) were taken instead of the sequence (H, Q). The eigenvalues $\{\lambda_i\}$ are the MS signal to noise ratios for the measurements defined by (H_{EU}, I) where the signal is the prior estimation error for each parameter measured by H_{EU} and the MS noise is 1.0. The eigenvalues depend on the nature of the prior information, as given by P_o, as well as on the sequence being processed. If $\lambda_i \gg 1$ then the measurement is a good one relative to the signal and good performance is obtained for the parameter defined by the corresponding column of H_{EU}. The converse occurs for those parameters corresponding to small eigenvalues. If the eigenvalues range from

16

very small to very large then computational difficulties may occur in processing the sequence. Such difficulties appear to be inherent in many problems of practical interest, especially in orbit determination. An interesting method of alleviating such difficulties based on the size of eigenvalues is reported in reference 12.

Influence of the initial covariance on estimation- Equation (5) is used for sequential estimation, beginning with the initial conditions $\hat{x}_A$, P_A. For linear systems obtained as approximations of nonlinear ones, it is assumed that $\hat{x}_A$ is zero and P_A is, in principle, determined from the measurements used to compute the reference solution of the nonlinear system. However, P_A is often guessed rather than calculated from the initial measurements. This can be done if the final estimate is approximately independent of the initial covariance, which occurs when the information from subsequent measurements is much greater than the initial information.

The posterior information and estimate from processing (H, Q, y) are (from eq. (5))

$$P^{-1} = P_A^{-1} + HQ^{-1}H^T$$

$$\hat{x} = PHQ^{-1}y$$

The condition, $\hat{x}_A = 0$, is used in the second equation above. If the information from (H, Q) is much greater than the initial information

$$HQ^{-1}H^T \gg P_A^{-1}$$

Then P_A has negligible influence on the estimate

$$\hat{x} \cong (HQ^{-1}H^T)^{-1}Q^{-1}y = H_E^{-T}y_E$$

where (H_E, I, y_E) is equivalent to (H, Q, y). This requires that the whole space be observable to the measurements, H, and if P_A^{-1} is expressed in the form

$$P_A^{-1} = H_E(MM^T)H_E^T$$

MM^T has eigenvalues much less than 1.0.

More generally, if the whole space is not observable then the initial information can be separated by selecting H_m, H_u to be any bases, respectively, of the observable and unobservable subspaces for the measurements, H

$$P_A^{-1} = H_m[H_m^T P_A H_m]^{-1}H_m^T + I_u$$

where

$$I_u = H_u [H_u^T P_A H_u]^{-1} H_u^T$$

The two parts of P_A^{-1} are the initial information matrices associated, respectively, with $\mathcal{H}_m$ and $\mathcal{H}_u$ and they are independent of the particular bases chosen for the construction. When H_m is the uncorrelated basis for (H, Q) and P_A, the posterior covariance becomes

$$P^{-1} = H_{EU}[I + \Lambda^{-1}] H_{EU}^T + I_u$$

If the eigenvalues satisfy the condition that

$$\lambda_i \gg 1 \qquad i = 1, \ldots, r*$$

then

$$P^{-1} \cong H_{EU} H_{EU}^T + I_u = HQ^{-1} H^T + I_u$$

$$\hat{x} \cong [I_u + HQ^{-1} H^T]^{-1} HQ^{-1} y$$

The estimation is approximately independent of the initial information associated with $\mathcal{H}_m$ but does depend on initial information associated with $\mathcal{H}_u$. Nevertheless, the estimate, $h^T \hat{x}$, of any parameter in $\mathcal{H}_m$ is approximately independent of all initial information. This can be verified if we let H_m, H_u be, respectively, any equivalent primary bases of (H, Q) and any square root of I_u, so that

$$P^{-1} = HQ^{-1} H^T + I_u = \left[H_m \mid H_u \right] \left[H_m \mid H_u \right]^T$$

Since both h and the measurements H are in $\mathcal{H}_m$, they can be given uniquely in terms of H_m in the form

$$h = H_m \bar{a} \qquad H = H_m M$$

whence

$$h^T \hat{x} = h^T PHQ^{-1} y \cong \bar{a}^T MQ^{-1} y$$

This result is approximately valid for all h in $\mathcal{H}_m$ and is independent of P_A. Note that the product, $\bar{a}^T M$, is independent of the choice of equivalent primary basis, H_m, used in the construction.

18

Secondly, if only some of the eigenvalues are much greater than 1.0 then the estimate, $h^T\hat{x}$, is independent of P_A only for those parameters for which h is in the subspace spanned by the columns of H_{EU} corresponding to these eigenvalues.

Finally, this discussion assumes that P_A is "reasonably" guessed so that at least the orders of magnitude of eigenvalues are correct. If some eigenvalues are small or a subspace is unobservable to the measurements, then errors in the guessed P_A will affect the estimate.

Selective Information Processing

A given measurement sequence has many equivalent bases and one can be found such that some of its measurements are in any specified subspace of $\mathcal{H}$ while the remaining are outside that subspace. Information consisting of measurements of parameters in the specified subspace can therefore be separated from the original sequence. This procedure can also be used to separate the processing equations, (5), into independent lower order parts by extracting and deleting certain information.

[11] Information separation. Let (H, Q) be any given measurement sequence and $\mathcal{H}_A$ any given subspace of $\mathcal{H}$, and let

$$\mathcal{H}_m = \mathcal{L}\{H\}$$

$$k = \dim(\mathcal{H}_A \cap \mathcal{H}_m)$$

An equivalent basis sequence of (H, Q) exists having k columns which are a basis of $\mathcal{H}_A \cap \mathcal{H}_m$.

The proof of [11] in appendix B outlines the computation of the required equivalent basis; that is, of a partitioned primary basis

$$\left(\left[H_{AM} \mid H_R\right], \quad I\right) \doteq (H, Q)$$

in which the k columns of H_{AM} form a basis of $\mathcal{H}_A \cap \mathcal{H}_m$ while the remaining $r^* - k$ columns, H_R, are not in $\mathcal{H}_A$ (r^* is the rank of H). The partitioned equivalent basis generated from [11] separates the information matrix into the sum

$$HQ^{-1}H^T = H_{AM}H_{AM}^T + H_R H_R^T$$

The two parts of the sum are unique and have rank k and $r^* - k$, respectively.

The capacity of the original sequence to reduce the estimation errors of parameters in $\mathcal{H}_A$ is not generally isolated in the extracted measurements, H_{AM}. This is seen as follows. The uncorrelated complement space of $\mathcal{H}_A$ is given by

$$\mathcal{H}_B = \{h: \quad h^T P_o H_A = 0\}$$

where H_A is any basis of $\mathcal{H}_A$. The space $\mathcal{H}_B$ is the set of costates of all parameters whose prior estimation errors are uncorrelated with the prior estimation errors of every parameter with costate in $\mathcal{H}_A$. The columns of H_{AM} are in $\mathcal{H}_A$ and, if processed, contribute nontrivial error reduction for parameters with costates in $\mathcal{H}_A$. The columns of H_R are not in $\mathcal{H}_A$ but, if processed, also reduce errors for parameters whose costates are in $\mathcal{H}_A$ except in the special case that the columns of H_R are in the complement space, $\mathcal{H}_B$.

On the other hand, error reduction in $\mathcal{H}_B$ can be prevented by extracting and processing only the measurements, H_{AM}, since these are all in the complement space of $\mathcal{H}_B$. If this is done, then the posterior estimation errors for parameters whose costates are in $\mathcal{H}_A$ will remain mutually uncorrelated with those whose costates are in $\mathcal{H}_B$.

The preceding remark can be elaborated to separate the processing equation into two independent lower order parts associated with $\mathcal{H}_A$ and $\mathcal{H}_B$. This is done by a two-part separation of (H, Q) into

$$(H, \ Q) = \begin{bmatrix} H_{AM} & \vdots & H_R \end{bmatrix}$$

Then H_R is further separated into

$$H_R = \begin{bmatrix} H_{RB} & \vdots & H_R' \end{bmatrix}$$

where H_{RB} is a basis of $\mathcal{L}\{H_R\} \cap \mathcal{H}_B$ and H_R' is the final remainder with columns outside both $\mathcal{H}_A$ and $\mathcal{H}_B$. Thus, an equivalent basis of the form

$$(H, \ Q) = \left(\begin{bmatrix} H_{AM} & \vdots & H_{RB} & \vdots & H_R' \end{bmatrix}, \ I \right)$$

is obtained. The subspace $\mathcal{H}_A$ is unobservable to the measurements H_{RB}, and $\mathcal{H}_B$ is unobservable to H_{AM}. The remaining measurements H_R' contain that information which causes the posterior estimation errors of parameters with costates in $\mathcal{H}_A$ to become correlated with those of parameters having costates in $\mathcal{H}_B$. Consequently, if H_R' is deleted from the data processing then the calculations (eq. (5)) can be separated into two independent parts of order dim$(\mathcal{H}_A)$ and n-dim$(\mathcal{H}_A)$ which are associated, respectively, with the estimation of parameters having costates in $\mathcal{H}_A$ from the measurements H_{AM} and with the estimation of parameters having costates in $\mathcal{H}_B$ from the measurements H_{RB} (see appendix B). This separation is maintained if a series of measurement sequences is separated and processed in the same way.

Discussion

Equivalent sequences and data are used here to obtain a reasonably unified view of many operations that can be applied to the processing of arbitrary finite sets of measurements. Equivalent sequences and data are any set of measurements and data (fictitious) which, when processed, give the same state estimate and covariance as the real sequence and data. An equivalent basis sequence is one which has the fewest number of measurements possible for an equivalent sequence, this number being the rank of the measurement costates of the real sequence. Basis sequences can be determined using statements [4] and [5] and equation (12). Several equivalent basis sequences with special properties were derived (e.g., the uncorrelated equivalent basis which can be used to describe performance, and bases which separate the information into measurements which are either in or outside of any selected subspace of the costate space).

Practical applications of these analytical results are not suggested as this matter is beyond the scope of this report and success is difficult to predict without computational experimentation. Computational efficiency might be improved in some problems by compressing the obtained information to an equivalent basis sequence prior to processing in equation (5). Equivalent bases similar to the uncorrelated basis are already known in the literature and applications to treating computationally ill-conditioned problems have been suggested (ref. 12). Finally, the information separation procedure can be used when only a few parameters are to be estimated and sufficient information on these parameters can be separated from the original sequence and processed. This procedure can also be used to separate the data processing equations into independent lower order parts which might prove useful in problems that separate naturally into nearly uncoupled parts.

In this section the data were analyzed in a context that is, perhaps, too general to yield much insight into the nature of estimation. In particular, the measurement sequences are assumed to be arbitrary, but, in practice, measurements are often obtained by sampling one or more data types at some arbitrary sequence of times. The data type, given as a time-dependent function of the state, is then a constraint on the measurements, and its characteristics are common to any sampling of the data type. For example, some states may be unobservable to the data types and, therefore, to any sampling. In the next section some elementary aspects of the analysis of observability for data types are considered.

DATA TYPES

Measurements considered in this section are restricted to samplings of one or more known scalar outputs or data types, of the form of equation (3), at some arbitrary finite sequence of times in an interval $[t_0, t_f]$. It is shown that the state and parameter spaces can be decomposed into subspaces termed observable and unobservable to these specific data types in a manner analogous to that given earlier for arbitrary measurement sequences. The

states which are unobservable to the data types are also unobservable to any sampling of the data types. If unobservable states occur the estimation equation (5) can be reduced in order by the dimension of the unobservable subspace since the problem can be reduced to one of estimating a lower order observable subsystem.

Observability Characteristics of Data Types

The measurements processed with equation (5) are restricted to samplings of one or more scalar data types of the form

$$m(t) = h(t)^T x(t_o) \qquad t_o \leqslant t \leqslant t_F$$

where the measurement costate $\{h(t), \; t_o \leqslant t \leqslant t_F\}$ is a known, continuous vector function. If more than one data type is available for sampling, then additional vector functions $\{h_i(t), \; i = 1, \; . \; . \; ., \; k\}$ are given.

The system is said to be completely observable on $[t_o, \; t_F]$ provided the state can be determined from the (noise-free) output $\{m(t), \; t_o \leqslant t \leqslant t_F\}$ (ref. 7). The system is completely observable on $[t_o, \; t_F]$ if, and only if, the integral

$$M(t_o, t_F) = \int_{t_o}^{t_F} h(t) h^T(t) dt$$

is nonsingular or alternatively, if there exists at least one set of n times in $[t_o, \; t_F]$ such that the matrix of measurement costates

$$[h(t_1), \; . \; . \; ., \; h(t_n)]$$

is nonsingular.

The elementary definitions and conditions for an observable system are somewhat more extensive than the above summary (e.g., refs. 1, 7, 8), but for present purposes the summary suffices. In many practical problems computations to determine observability are difficult but this subject is beyond the scope of this paper.

The concern here is with systems that may or may not be observable. If the data types $\{h_i(t), \; i = 1, \; . \; . \; ., \; k\}$ are available for sampling then each data type spans a subspace of $\mathcal{H}$

$$\mathcal{H}_{m_i} = \mathcal{L}\{h_i(t), \; t_o \leqslant t \leqslant t_F\} \tag{18}$$

called its observable parameter space. A basis of this space is given by the row or column space of the matrix

$$M_i(t_o, t_F) = \int_{t_o}^{t_F} h_i(t) h_i^T(t) dt , \qquad i = 1, \ldots, k \qquad (19)$$

The observable parameter space for the collection of data types is the direct sum

$$\mathcal{H}_m = \mathcal{H}_{m_1} \oplus \mathcal{H}_{m_2} \oplus \ldots \oplus \mathcal{H}_{mk} \qquad (20)$$

which is the space spanned by the collection of bases computed from equation (19). If $\mathcal{H}_m$ is r-dimensional, then there exists a set of r times in $[t_o, t_F]$ and corresponding data types such that the corresponding measurement costates are a basis of $\mathcal{H}_m$. Any other sequence spans a space included in $\mathcal{H}_m$.

The unobservable state space is defined from the data types by

$$X_u = \{x: \quad x^T h_i(t) = 0; \ t_o \leqslant t \leqslant t_F; \ i = 1, \ldots, k\} \qquad (21a)$$

If a basis, H_m, of $\mathcal{H}_m$ is given then X_u can be determined from

$$X_u = \{x: \quad x^T H_m = 0\} \qquad (21b)$$

An orthogonal decomposition of $\mathcal{H}$ and X is completed by defining

$$\mathcal{H}_u = \{h: \quad h^T P_A H_m = 0\} \qquad (22a)$$

$$X_m = \mathcal{L}\{P_A H_m\} \qquad (22b)$$

Here, P_A refers to the initial covariance used in the sequential processing. The matrix, P_A, results from some sufficient set of initial measurements of unspecified type, but all subsequent measurements are assumed samplings of the data types. The subspaces defined in equations (22) are based partly on the initial statistical properties of the estimation errors as given by P_A, which specifies those parameters whose initial estimation errors are correlated with the initial estimation error of any parameter that can subsequently be measured by sampling the data types.

The subspaces defined by equations (20) to (22) are analogous to those given by (13) for arbitrary measurement sequences. It is readily verified that

$$\mathcal{H} = \mathcal{H}_m \oplus \mathcal{H}_u \text{ and } \mathcal{H}_m \cap \mathcal{H}_u = 0$$

so that any costate, h, can be given uniquely as a sum

$$h = h_m + h_u$$

23

whose parts are, respectively, in $\mathcal{H}_m$ and $\mathcal{H}_u$. Similar statements apply to χ, χ_m, χ_u and any state x. In addition

$$\dim[\mathcal{H}_m] = \dim[\chi_m] = r$$

$$\dim[\mathcal{H}_u] = \dim[\chi_u] = n - r$$

where n is the dimension of the state space. These subspaces are orthogonal decompositions of $\mathcal{H}$ and χ in the sense of the inner products

$$\langle h_1, h_2 \rangle = h_1^T P_A h_2 \text{ and } \langle x_1, x_2 \rangle = x_1^T P_A^{-1} x_2$$

Finally, related bases of the subspaces are given by

$$\chi_m = P_A H_m \text{ and } \chi_u = P_A H_u \tag{23}$$

where H_m and H_u are any bases of $\mathcal{H}_m$ and $\mathcal{H}_u$, respectively.

Reduction to an Observable Subsystem

If unobservable states occur then a suitable transformation of the state space gives the data types in terms of a reduced set of new variables and the estimation equations (5) are separated into trivial and nontrivial lower order equations. Only the nontrivial equations need be carried in the computations. To accomplish this, define the partitioned transformation of the state space

$$\chi = \left[\chi_m \vdots \chi_u \right] \tag{24}$$

in which χ_m (r columns) and χ_u ($n - r$ columns) are, respectively, bases of the subspaces χ_m and χ_u given by equations (21) and (22). Any state can be given as a linear combination of the columns of χ

$$x = \left[\chi_m \vdots \chi_u \right] \begin{pmatrix} w_m \\ w_u \end{pmatrix} = \chi_m w_m + \chi_u w_u \tag{25}$$

This gives x as the sum of two vectors, one each in χ_m and χ_u. The components of w_m ($r \times 1$ vector) and w_u ($n - r \times 1$ vector) are the components of x in χ_m and χ_u, respectively, for the particular bases of these subspaces chosen for the transformation, (24).

The costate of any measurement permitted by the data types is in $\mathcal{H}_m$; hence, measurements can be given as

$$y = h^T x + \varepsilon = h^T X_m w_m + \varepsilon$$

so that measurements can be given as linear combinations of the r new variables of w_m and the linear combination is defined by the $r \times 1$ vector, $X_m^T h$. Consequently, any measurement sequence and its data (H, Q, y) can be given as (M, Q, y), where

$$M = X_m^T H \qquad (26)$$

The initial covariance of $\tilde{w}$ is obtained from the transformation

$$W_A = X^{-1} P_A X^{-T} = \begin{bmatrix} (X_m^T P_A^{-1} X_m)^{-1} & 0 \\ 0 & (X_u^T P_A^{-1} X_u)^{-1} \end{bmatrix} \qquad (27)$$

The off-diagonal submatrices are zero since $X_m^T P_A^{-1} X_u = 0$; that is, the two subspaces are orthogonal. The nontrivial parts of W_A are the initial covariances of $\tilde{w}_m$ and $\tilde{w}_u$ and are represented hereafter as W_{mA} and W_{uA}. If any sequence of measurements and its data (H, Q, y) are now processed and the resulting posterior covariance P is transformed, one obtains the posterior covariance of $\tilde{w}$ as

$$W = \begin{bmatrix} (W_{mA}^{-1} + M Q^{-1} M^T)^{-1} & 0 \\ 0 & W_{uA} \end{bmatrix}$$

and the estimate $\hat{w}$ is given by

$$\hat{w}_m = W_{mA} M [Q + M^T W_{mA} M]^{-1} y$$

$$\hat{w}_u = 0$$

Finally, if subsequent batches of measurements are processed sequentially then the transformation, (24), maintains the separation of the processing equations at every step

$$\left.\begin{aligned} W_m &= W_{mo} - W_{mo} M [Q + M^T W_{mo} M]^{-1} M^T W_{mo} \\ \hat{w}_m &= W_{mo} M [Q + M^T W_{mo} M]^{-1} (y - M^T \hat{w}_m^-) + \hat{w}_m^- \end{aligned}\right\} \qquad (28)$$

$$\left.\begin{aligned} W_u &= W_{uA} \\ \hat{w}_u &= 0 \end{aligned}\right\} \qquad (29)$$

Here, W_{m0} and $\hat{w}_m^-$ are the prior covariance of $\tilde{w}_m$ and estimate of w_m obtained from processing all earlier measurements, and (M, Q, y) is the current batch transformed by equation (26). Only equation (28) describing the r variables w_m need be carried in the computations. These are entirely analogous to the original processing equations, (5), and are smaller in order than (5) by the dimension of $\mathcal{H}_u$.

It is usually more convenient to begin with bases of $\mathcal{H}_m$ and $\mathcal{H}_u$ and then obtain X_m and X_u from these bases by using equation (23). If this is done, then the new variables w_m, w_u can be written

$$w_m = (H_m^T P_A H_m)^{-1} H_m^T x$$

$$w_u = (H_u^T P_A H_u)^{-1} H_u^T x$$

that is, the new variables are parameters with costates in $\mathcal{H}_m$ and $\mathcal{H}_u$. The basis, H_m, can always be selected so that $H_m^T P_A H_m$ is the unit $r \times r$ matrix. If this is done then

$$W_{mA} = I$$

and M can be computed from

$$M = H_m^\# H$$

(This result is established by noting that $M = H_m^T P_A H = (H_m^T P_A H_m) H_m^\# H = H_m^\# H$.) Thus, to carry out the processing with the reduced equations (28), it is necessary to calculate one matrix, $H_m^\#$, which is then applied to the costates of each measurement sequence, H, subsequently processed.

The covariance and estimate of the original state variables, x, is readily recovered from W, $\hat{w}$ at any step

$$\left. \begin{aligned} P &= XWX^T = X_m W_m X_m^T + X_u W_{uA} X_u^T \\[2ex] \hat{x} &= X\hat{w} = X_m \hat{w}_m \end{aligned} \right\} \tag{30}$$

The estimated state is always an element of X_m that is, is an observable state. The covariance is separated in equations (30) into two parts, this separation being independent of the particular bases of X_m, X_u used in the transformation. The covariance of $\tilde{x}_u$, given by

$$P_u = X_u W_{uA} X_u^T$$

is unreduced from its initial value by processing any measurements obtained from the data types. Therefore, P_u is a lower bound on the MS estimation errors and, in particular, its diagonal elements are lower bounds on the MS errors to which the original state variables can be estimated from the data types.

A potential source of progressive accuracy loss in calculations is apparent in (30). Assume that χ_u is nontrivial and that the estimation calculations are not separated as in (28) and (29). As data are then processed, the matrix W_m in equations (30) is reduced while P_u remains fixed. This can result in the progressive loss of numerical significance and even the rank with which $X_m W_m X_m^T$ is retained in the calculated covariance matrix. The estimate, $\hat{x}$, depends on $X_m W_m X_m^T$ and not on P_u making it subject to progressively larger calculation errors. The removal of unobservable states from the data processing, as in equation (28), removes this source of calculation errors.

An example is given in appendix D to illustrate the application of observability analysis described in this section.

Discussion

The data type refers to a scalar function of time and the initial state whose sampling at some arbitrary finite set of times in an interval $[t_o, t_F]$ provides the measurement sequence used to estimate the state. The data type is then a constraint on the measurements and all samplings possess common properties as a result of the constraint.

One such property is observability which refers to the sufficiency of the data type for determination of the state from some sampling in the interval $[t_o, t_F]$. The observable parameters and unobservable states are defined from the data type. The initial covariance matrix is introduced and then the unobservable parameters and observable states can be defined. These four subspaces define an orthogonal decomposition of the parameter and state spaces and permit separation of the processing equations into trivial and nontrivial parts associated with estimating the observable and unobservable components of the state. If the system is not observable to the data type, the problem can be reduced to one of estimating the lower order observable subsystem.

This section has considered only basic definitions and the immediate effects of unobservable states in the data processing; it is by no means as extensive as one expects the analysis of data types eventually to be. It is possible to distinguish between complete and total observability (ref. 7). The observable space, $\mathcal{H}_m$, is completely observable in $[t_o, t_F]$ if it is spanned by the data type during this interval, and is totally observable if it is spanned during every nontrivial subinterval of $[t_o, t_F]$. Apparently, the space spanned by the data type during $[t_o, t_F]$ is totally observable, provided the data type can be expanded in a Taylor series on $[t_o, t_F]$, and this is often the case in estimation problems. In addition, only the integral condition for observability has been mentioned, but conditions in terms of derivatives of the data type costate are also available, by way of the duality

between controllability and observability, from the results in reference 13 and elsewhere. Computations to determine the space observable to a data type are difficult at best in many practical problems, but the matter of computationally efficient methods is beyond the scope of this report. Finally, the characteristics of the data type which influence performance when all states are observable are not investigated here.

CALIBRATION ACCURACY LIMITS IN ESTIMATION PROBLEMS

In most estimation problems, calibration measurements are carried out initially and later measurements are comparisons between unknowns in the problem and calibrated quantities. It is expected that the accuracy to which the unknowns can be estimated from such measurements is limited by the accuracy of the initial calibrations. The existence of accuracy limits in an estimation problem implies the existence of corresponding unobservable states and parameters and suggests that observability theory can be applied to their analysis, as is done below in the case of the generally occurring calibration accuracy limits.

Measurements are comparisons of like physical phenomena, such as two lengths or two masses. The basis of comparison is made unique by international definition of unit phenomena of time, length, mass, and temperature termed the second, meter, kilogram, and degree Celsius.[2] For example, the meter is currently defined to be the distance covered by 1,650,763.73 wavelengths of a certain line of a Krypton 86 source under specified conditions (ref. 14). Estimation problems can be regarded as exercises in comparing some set of physical phenomena to these internationally defined unit phenomena. *Calibrations* are defined here as comparisons of any quantity in the problem with the unit phenomena (e.g., measuring a distance by counting wavelengths). With the exception of time, which is sometimes measured with a cesium clock, the calibrations are usually carried out initially and are excluded from the data types. In these cases, the accuracy to which the state variables can be estimated in the international units by sampling the data types is necessarily limited to the accuracy set by the initial calibrations.

The nature of the calibration accuracy limit is obvious in simple cases. For example, a length estimated by comparison with a meter stick can eventually be estimated, after a series of such independent measurements, with nearly the accuracy with which the meter stick was originally calibrated. If the length is compared to five independently calibrated meter sticks then the unknown length and the lengths of each of the five meter sticks can be estimated to the accuracy of the best initial estimate of length that here, is the length of all five meter sticks laid end to end, or equivalently, the accuracy obtained by averaging five independent calibrations of one meter stick.

---

[2]Other units are derived from these four (ref. 14). In orbit estimation only time and length units appear.

28

In both cases, sampling the data types amounts to a comparison of all lengths in the problem with an intermediate length standard, the length of the meter stick in the first case and the sum of the lengths of the five meter sticks in the second. The estimation accuracy for the intermediate length standard cannot be improved by sampling the data type. It is unobservable.

Other problems are less obvious since calibrations of varying accuracies are often buried in the initial estimates of fundamental parameters in the problem. However, features similar to those described above can be demonstrated; if the data types are not calibrations, then an independent unobservable state can be shown to exist for each physical unit in the problem. For each such state a corresponding accuracy limit is obtained, these limits being the best accuracies to which any length, mass, and temperature in the problem have been estimated initially. The length, mass, and temperature which have been estimated initially to these best accuracies are the intermediate standards of length, mass, and temperature. They are the "meter sticks" of the problem against which the data types compare other quantities.

Unobservable Calibration States

If the data types are not calibrations, they are comparisons or ratios of unknown quantities in the problem and, hence, are nonlinear in the unknowns. Therefore, for the analysis of calibration accuracy the linear system (eqs. (1) to (3)) is assumed to be obtained by linearizing a nonlinear one (eq. (6)), using equations (8) and (9) and a reference solution of (6a) given from some set of initial measurements which include the calibrations.

Sufficient conditions for existence of unobservable states- It is shown, next, that simple conditions on the nonlinear system suffice for the existence of unobservable states of the linearized system. These results are applied in the subsequent analysis of calibration accuracy.

In the following statement, X is the n × 1 state vector, and $\{X_A(t), \ t_0 \leq t \leq t_F\}$ is the reference solution of the nonlinear state equation used in linearizing the equations. The data types, $\{m_i(X), \ i = 1, \ . \ . \ ., \ k\}$ and the components of the state equation

$$\dot{X}(t) = \bar{f}(X(t))$$

are all assumed defined with continuous partial derivatives on the state space.

[12] Sufficient conditions. Define the diagonal matrices

$$J = \begin{bmatrix} \ddots & & \\ & J_i & \\ & & \ddots \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \ddots & & \\ & \lambda^{J_i} & \\ & & \ddots \end{bmatrix}$$

where $\{J_1, \ . \ . \ ., \ J_n\}$ is a set of integers and λ is real and arbitrary. If there exists a set of integers, $\{J_1, \ . \ . \ ., \ J_n\}$, for which

$$(1) \quad \bar{f}(\Lambda X) = \Lambda \bar{f}(X)$$

$$(2) \quad m_i(\Lambda X) = m_i(X) \qquad i = 1, \ldots, k$$

for all X and λ, then the state

$$\xi_0 \doteq J\, X_A(t_0)$$

is an unobservable (deviation) state of the linearized system.

A detailed proof of statement [12] is given in appendix C. The conditions of the statement require the data types and the components of $\bar{f}(X)$ to be algebraically homogeneous. As will be shown, these conditions are met if the data types exclude calibrations. Then the integer powers to which any single physical unit, other than the unit of time, appears in the state variables are sets of integers that satisfy the conditions of statement [12]. These integers are readily given by inspection of the units of the state variables. As shown in appendix C, it follows from Euler's theorem for homogeneous functions that the consequence of the first condition of statement [12] is that the vector function

$$\xi(t) \equiv J\, X_A(t) \qquad t_0 \leqslant t \leqslant t_F \tag{31}$$

is a solution of the linearized state equation and can be given as

$$\xi(t) = \phi(t, t_0) J\, X_A(t_0) = \phi(t, t_0)\xi_0$$

The consequence of the second condition is that the data types satisfy

$$\xi(t)^T \nabla m_i(t) = \xi_0^T \phi^T(t, t_0)\nabla m_i(t) = 0$$

Noting equation (21a) and equations (3) and (9), then ξ_0 is an unobservable state of the linearized system.

Similar consequences occur for the nonlinear system as well. If the first condition is satisfied by some set of integers, then the vector function

$$\psi(t) \equiv \Lambda X_A(t) \qquad t_0 \leqslant t \leqslant t_F \tag{32}$$

is a solution of the nonlinear state equation for all λ. If, in addition, the second condition is satisfied by the same integers, then these solutions are all indistinguishable to the data types; that is,

$$m_i(\Psi(t)) = m_i(X_A(t)) \qquad i = 1, \ldots, k$$

and the output is identical for all solutions defined by (32).

State equation and condition one- It is next shown that the state equation satisfies the first condition of statement [12] for those sets of integers which are the integer powers to which any single physical unit (meter, kilogram, degree), other than the unit of time, appears in the state variables.

First, a restriction on the definition of the state is necessary. The state, roughly, is a vector list of independent unknowns which suffice to define the measurements and the observed system. Every physical quantity in the problem is assumed estimated initially, as is implied by the existence of a reference solution of the state equation. However, the true value of every such quantity is unknown and any difference between the true and estimated values can affect the data. Only the international unit phenomena, whose values are one by definition, are regarded here as known. In many practical problems, some quantities, which are estimated very accurately initially, are treated as known, but such an approximation cannot be made here without eliminating the calibration accuracy limit. In addition, the state variables are assumed independent (no state variable can be given as a function of the others) and suffice to define the measurements and the observed system (all unknowns in the problem are state variables or functions of state variables).

The components of the state vector are given in units of seconds, meters, kilograms, or degrees Celsius, which appear to arbitrary integer powers (positive, negative, or zero)

$$[X_i] = \sec^{\alpha_i} m^{\beta_i} kgm^{\gamma_i} \deg^{\delta_i} \qquad i = 1, \ldots, n \qquad (33)$$

If one of the units, say length, is scaled by λ

$$m = \lambda m'$$

where λ is any real number,[3] then the same state expressed in the two units is related by

$$X' = \left[\begin{array}{c} \lambda^{\beta_i} \end{array} \right] X = \Lambda_L X \qquad (34)$$

If $\{X_A(t), X'_A(t), t_o \leq t \leq t_F\}$ are the same solution of the state equation viewed in the two units, both satisfy the state equation

$$\dot{X}_A(t) = \bar{f}(X_A(t)) , \qquad \dot{X}'_A(t) = \bar{f}(X'_A(t))$$

---

[3]The new unit, m', defines a different number of wavelengths of krypton 86 to be the distance unit. This changes the numerical size, or scale, of all distances by the factor λ. No state variable can become the distance, or known unit, by such a change of scale.

The vector function, $\bar{f}(\ )$, is unaffected by changes in the scale of any unit since all quantities requiring units are necessarily unknowns and are state variables or functions of state variables by definition. From this and equation (34) it follows that

$$\bar{f}(\Lambda_L X_A(t)) = \Lambda_L \bar{f}(X_A(t)) \tag{35}$$

where

$$\Lambda_L = \left[\begin{array}{ccc} & & \\ & \lambda^{\beta_i} & \\ & & \end{array} \right]$$

Thus, the state equation satisfies statement [12] for the integers $\{\beta_i, i = 1, \ldots, n\}$ which are the powers of the length unit in the state variables. The same is true for the sets of integers, $\{\gamma_i, i = 1, \ldots, n\}$, etc., corresponding to any other unit, except time, which appear in the state variables.

Data types and condition two- In this section it is shown that the data types satisfy condition two of statement [12] for those integers given above with equation (35), provided only that the data types exclude calibrations.

Measurements are quantitative comparisons of like physical phenomena. For example, radar ranging stations measure the phase shift between emitted and received waves, this measurement being a relation between the observed system and the station given by

$$m(X) = \frac{1}{2\pi} \frac{|\bar{R}_V - \bar{R}_S|}{\lambda_S}$$

which is a comparison of the station-vehicle distance, $|\bar{R}_V - \bar{R}_S|$, with the radar wavelength, λ_S. All physical quantities required to define this data type $\bar{R}_V$, $\bar{R}_S$, λ_S appear among or are functions of the state variables by definition of the state.

The basis for comparing like phenomena is made unique by defining a single such phenomenon to be the unit. Calibrations are defined here to be comparisons of physical quantities with the appropriate unit phenomenon. The values obtained from calibrations are given in units: seconds, meters, kilograms, or degrees Celsius. It is assumed that the state is to be estimated in these units so that calibrations must occur somewhere in the estimation procedure. Any other measurement is a ratio of like phenomena, neither of which is a unit. Such ratios are nonlinear in the unknowns and their values are dimensionless and, therefore, unaffected by any change in the size of units.

Assume that calibrations occur in an initial set of measurements. Then the data types, $\{m_i(X), i = 1, \ldots, k\}$, to be sampled are dimensionless ratios of functions of the state variables whose values are unaffected by

changes in the size of units. If X, X' refer to the same state given in each of two sets of units that differ in size, then

$$m_i(X') = m_i(X) \qquad i = 1, \ldots, k$$

where X, X' are related by (34) for changes in the length scale, and by analogous relations for scale changes in any other unit. Therefore

$$m_i(\Lambda X) = m_i(X) \qquad i = 1, \ldots, k \qquad (36)$$

$$\Lambda = \begin{bmatrix} \lambda^{J_i} \end{bmatrix}$$

where $\{J_i\}$ are the integer powers to which any single unit appears in the state variables and λ is any real number. Thus the data types satisfy condition two of statement [12] for the integers given above, provided the data types are not calibrations. In view of equations (35) and (36), it can be stated that:

[13] Statement [12] is satisfied by those sets of integers $\{J_i, i = 1, \ldots, n\}$, which are the integer powers to which any single physical unit, other than the unit of time, appears in the state variables provided the data types are not calibrations.

Calibration Accuracy, the Meter Sticks, and Their
Relation to Initial Measurements

If the data types are not calibrations then, consequent to statements [12] and [13], unobservable states occur and are defined by

$$\xi_O = J X_O, \qquad J = \begin{bmatrix} J_i \end{bmatrix} \qquad (37)$$

where X_O is the initial state of the reference trajectory and $\{J_1, \ldots, J_n\}$, are the integer powers to which any single unit, other than the unit of time, appears in the state variables. For example, the integers $\{J_1, \ldots, J_n\}$ for the length unit are the powers to which the meter is raised in the components $X_1, \ldots, X_n$ of the state vector. For each unit (length, mass, or temperature) which appears in the state variables, one independent unobservable state $\{\xi_{O\ell}, \xi_{Om}, \xi_{O\tau}\}$ is defined by (37).

The initial state of the reference trajectory is assumed to be an estimate based on some arbitrary and sufficient set of initial measurements. It is also assumed that the error covariance matrix, P_A, corresponding to this estimate is given and is a positive definite symmetric $n \times n$ matrix. A basis of the unobservable calibration states is

33

$$X_u = \left[\xi_{o\ell} \xi_{om} \xi_{o\tau} \right] \qquad (38a)$$

and a basis of its complement space is defined by

$$X_m = \text{any basis of } \{X: \quad X^T P_A^{-1} X_u = 0\} \qquad (38b)$$

Earlier terminology is continued by referring to X_m as the observable states, but this is correct only in the absence of independent unobservable states other than (38a). The initial covariance can now be separated into parts associated with X_m and X_u:

$$P_A = X_m \left[X_m^T P_A^{-1} X_m \right]^{-1} X_m^T + X_u \left[X_u^T P_A^{-1} X_u \right]^{-1} X_u^T \qquad (38c)$$

or simply

$$P_A = P_m + P_u$$

It is also convenient to adopt the notation

$$\left[X_u^T P_A^{-1} X_u \right]^{-1} = \begin{bmatrix} \mu_{\ell\ell} & \mu_{\ell m} & \mu_{\ell\tau} \\ \mu_{\ell m} & \mu_{mm} & \mu_{m\tau} \\ \mu_{\ell\tau} & \mu_{m\tau} & \mu_{\tau\tau} \end{bmatrix}$$

in order to identify individual elements from this positive definite matrix.

Initial estimation accuracy for any length- A length in the problem is any differentiable scalar function of the state $\ell(X)$ which has dimensions of length. If the length unit is scaled by λ, then the value of $\ell(X)$ must also be scaled by λ to obtain $\ell(X')$ in the new unit, but if any other unit is scaled by λ, the value of $\ell(X)$ is unchanged:

$$\left. \begin{array}{l} \ell(\Lambda_\ell X) = \lambda\ell(X) \\ \ell(\Lambda_m X) = \ell(\Lambda_\tau X) = \ell(X) \end{array} \right\} \qquad (39)$$

where

$$\Lambda = \begin{bmatrix} \searrow \\ & \lambda^{J_i} \\ & & \searrow \end{bmatrix}$$

and the subscripts, l, m, τ, indicate that the integers $\{J_1, \ldots, J_n\}$ in Λ correspond to the length, mass, or temperature unit. From Euler's theorem (appendix C) for functions with the above homogeneity properties, it follows that

$$X_u^T \nabla l = \begin{pmatrix} l(X_0) \\ 0 \\ 0 \end{pmatrix} \tag{40a}$$

valid for any length, $l(X)$. The gradient is understood to be evaluated at X_0. Analogous statements follow for all functions, $m(X)$, $\tau(X)$ which are any mass, or temperature in the problem

$$X_u^T \nabla m = \begin{pmatrix} 0 \\ m(X_0) \\ 0 \end{pmatrix}, \qquad X_u^T \nabla \tau = \begin{pmatrix} 0 \\ 0 \\ \tau(X_0) \end{pmatrix} \tag{40b}$$

The *accuracy* to which any length is estimated from the initial measurements is its MS initial error ratio, that is,

$$\eta(l) \equiv \frac{\nabla l^T P_A \nabla l}{l^2(X_0)} = \frac{\nabla l^T P_m \nabla l}{l^2(X_0)} + \frac{\nabla l^T P_u \nabla l}{l^2(X_0)} \tag{41}$$

or simply

$$\eta(l) = \eta_m(l) + \eta_u(l)$$

The part, $\eta_u(l)$, is due to the unobservable initial errors associated with the calibration states. The accuracy to which $l(X)$ can be estimated from any subsequent processing of data from any noncalibration data types cannot exceed $\eta_u(l)$. Employing equation (40a) in the definition of P_u (eqs. (38)), obtain

$$\eta_u(l) = \mu_{ll} \tag{42a}$$

Thus, the accuracy limit is μ_{ll} and is the same for all lengths in the problem. In orbit estimation problems in which important lengths, such as radar station location, planetary distances or miss distance, are estimated from noncalibration data types, μ_{ll} is the limiting accuracy obtainable.

Analogous statements apply to all masses or temperatures in the problem

$$\eta_u(m) = \mu_{mm}, \qquad \eta_u(\tau) = \mu_{\tau\tau} \tag{42b}$$

More generally, quantities, $g(X)$, which have composite units (e.g., $m^\alpha kg^\beta deg^\gamma$) have the accuracy limit

$$\eta_u(g) = (\alpha, \beta, \gamma) \left[X_u^T P_A^{-1} X_u \right]^{-1} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \quad (42c)$$

Length with best initial accuracy- A length in the problem is any differentiable scalar function of the state $l(X)$ which has dimensions of length. The accuracy of the initial estimate of any such length is given by

$$\eta(l) = \eta_m(l) + \mu_{ll}$$

Since μ_{ll} is fixed and independent of $l(X)$ then the length which minimizes $\eta(l)$ is any function, $l^*(X)$, such that

$$\eta_m(l^*) = 0 \quad (43)$$

This requires

$$\nabla l^{*T} X_m = 0$$

By definition of X_m (eq. (38b)) all vectors which satisfy this equation must be linear combinations of the columns of $P_A^{-1} X_u$

$$\nabla l^* = P_A^{-1} X_u \bar{a}$$

Since $l^*(X)$ is a length it satisfies equation (40a) so that

$$\bar{a} = \left[X_u^T P_A^{-1} X_u \right]^{-1} \begin{pmatrix} l^*(X_o) \\ 0 \\ 0 \end{pmatrix}$$

and then

$$\frac{\nabla l^*}{l^*(X_o)} = P_A^{-1} \left(\xi_{ol} \mu_{ll} + \xi_{om} \mu_{lm} + \xi_{o\tau} \mu_{l\tau} \right) \quad (44a)$$

While equation (44a) does not give $l^*(X)$ uniquely, it does give the gradient of $l^*(X)$ at X_o within the ambiguous factor, $l^*(X_o)$. Analogous results are obtained for the mass or temperature with the best initial accuracies

$$\frac{\nabla m^*}{m^*(X_0)} = P_A^{-1}\left(\xi_{0l}\mu_{lm} + \xi_{0m}\mu_{mm} + \xi_{0\tau}\mu_{m\tau}\right) \tag{44b}$$

"Meter sticks" of the problem- In the absence of calibrations among the data types, the estimation of the state from the data types is equivalent to comparing all lengths in the problem to $l^*(X)$ and all masses to $m^*(X)$. The quantities, $l^*(X)$, $m^*(X)$, $\tau^*(X)$ are the intermediate standards or "meter sticks" of the problem since, for their respective physical phenomena, they occupy the role of a meter stick used to measure a distance. This is shown by the following analysis.

If the data types exclude calibrations then unobservable calibration states are routinely present and can be removed from the data processing as described in an earlier section. This is done by transforming the deviation state to a new set of variables, w, defined by

$$x = \begin{bmatrix} X_m & \vdots & X_u \end{bmatrix} \begin{pmatrix} w_m \\ --- \\ w_u \end{pmatrix} \tag{25}$$

In the present context, X_m and X_u are defined in equations (38). The new v~~ bles, w_m, w_u are the components of the state in X_m, X_u. The components, w_m, can be estimated from any (sufficient) noncalibration data types (eqs. (28)) while the components $\tilde{w}_u$ cannot (eqs. (29)).

The nature of the variables, $\tilde{w}_u$, can be determined by premultiplying equation (25) with $X_u^T P_A^{-1}$, and inverting the result to obtain

$$\tilde{w}_u = \begin{pmatrix} \nabla l^{*T}\tilde{\chi}/l^*(X_0) \\ \nabla m^{*T}\tilde{\chi}/m^*(X_0) \\ \nabla \tau^{*T}\tilde{\chi}/\tau^*(X_0) \end{pmatrix} \tag{45}$$

Thus, the quantities $\tilde{w}_u$ which cannot be estimated are the errors, $\tilde{l}^*$, $\tilde{m}^*$, $\tilde{\tau}^*$ with which the "meter sticks" of the problem have been estimated from the initial measurements.

Rather than analyze the nature of $\tilde{w}_m$ directly it is convenient to consider that part of the initial estimation error for any length, mass, or temperature in the problem which can be estimated from noncalibration data types. The estimation error, $\tilde{l}$, for any length, $l(X)$, is given from the state estimation error by

$$\tilde{l} = \nabla l^T \tilde{\chi}$$

This can be separated into observable and unobservable parts

$$\tilde{l} = \nabla l^T X_m \tilde{w}_m + \nabla l^T X_u \tilde{w}_u$$

or simply

$$\tilde{l} = \tilde{l}_m + \tilde{l}_u$$

The observable part is

$$\tilde{l}_m = \tilde{l} - \nabla l^T X_u \tilde{w}_u$$

which, noting equations (40a) and (45), becomes

$$\tilde{l}_m = \tilde{l} - \frac{l(X_0)}{l^*(X_0)} \nabla l^{*T} \tilde{x}$$

If the ratio of $l(X)$ to the meter stick is defined

$$r(X) = \frac{l(X)}{l^*(X)} \tag{46}$$

then the estimation error for this ratio is

$$\tilde{r}(X) = \nabla r^T \tilde{x} = \frac{1}{l^*(X_0)} \left[\tilde{l} - \frac{l(X_0)\tilde{l}^*}{l^*(X_0)} \right]$$

so that

$$\tilde{l}_m = l^*(X_0)\tilde{r} \tag{47}$$

Thus, for any length, $l(X)$, that part of the initial estimation error which can be estimated from samples of noncalibration data types is the error with which $l(X)$ is known relative to the meter stick, $l^*(X)$, of the problem. Similar conclusions apply to all masses and temperatures in the problem. The estimation procedure, when using noncalibration data types, is at most one of comparing all lengths to $l^*(X)$ and all masses to $m^*(X)$.

Calibration accuracy and initial measurements- The accuracy to which any quantity requiring units (other than time) can be estimated without performing new calibrations is limited. The relation of these limits to the measurements performed initially is examined next.

38

The initial measurements are those used to determine the reference trajectory. These have been loosely described as any arbitrary and sufficient set. It is further assumed that this set is finite; that a vector, Y, can be given which is a (finite) list of all independent physical quantities required to define the initial measurements; that the state variables, X, can be included in the variables of Y, and that the set of initial measurements is sufficient to determine Y.

In the preceding statement, the independence of the variables, Y, means that there is no nontrivial relation

$$g(Y) = 0$$

among these variables. Further, the vector, Y, can always be given as

$$Y = \begin{pmatrix} X \\ --- \\ Z \end{pmatrix} \qquad (48)$$

Here, X is the state vector which has appeared throughout the earlier text and is a list of independent quantities required to define the observed dynamic system and the data types. These can always be included in Y, since the reference initial state, X_0, could not otherwise be determined from the initial measurements. However, the initial measurements are not restricted to the data types so that additional unknown physical quantities, Z, are generally required for their definition. Finally, the sufficiency of the initial measurements to determine Y means that there is at least one subset of measurements for which an inverse function exists giving all the variables, Y, in terms of the data. This, of course, guarantees that X_0 can be determined.

The above description of the initial measurements is sufficient for immediate purposes but is incomplete. A statement to insure that all the initial measurements are relevant to the estimation of X_0 is lacking, as well as a statement on the measurement noise.

Next, list all the initial measurements and segregate them into calibrations of length, mass, and temperature and measurements that are not such calibrations; that is,

$$\ell_1(Y), \ell_2(Y), \ldots \qquad (49a)$$

refer to the initial length calibrations,

$$m_1(Y), m_2(Y), \ldots, \tau_1(Y), \tau_2(Y), \ldots \qquad (49b)$$

are the mass and temperature calibrations, and finally,

$$n_1(Y), n_2(Y), \ldots \qquad (49c)$$

39

are the remaining measurements. The measurements of this last category are dimensionless in length, mass, and temperature.

The measurements in equations (49) are required to be sufficient to determine Y; that is, some subset of (49) has an inverse function which gives Y in terms of the data. Let $H_A(Y)$ be the matrix whose columns are the gradients of equations (49) evaluated at Y, and let Q_A be the noise matrix. Take the requirement for the sufficiency of equations (49) to mean that $H_A(Y)$ has maximal rank at every Y, and, for simplicity, assume that all initial measurements are made with independent noise so that Q_A is diagonal. Then the initial measurements can be processed to obtain the minimum variance estimate Y_O and the initial covariance is given by

$$E[\tilde{y}\tilde{y}^T] = \left[H_A Q_A^{-1} H_A^T \right]^{-1} \tag{50}$$

where H_A is evaluated at Y_O and $\tilde{y}$ is the estimation error $Y - Y_O$.

The columns of H_A can be partitioned into two parts that contain the gradients of equations (49) with respect to X and Z, respectively, and these can be further partitioned according to the category of measurements

$$H_A = \begin{bmatrix} H_X \\ --- \\ H_Z \end{bmatrix} = \begin{bmatrix} H_{Xl} & H_{Xm} & H_{X\tau} & H_{Xn} \\ ---- & ---- & ---- & ---- \\ H_{Zl} & H_{Zm} & H_{Z\tau} & H_{Zn} \end{bmatrix} \tag{51}$$

The initial covariance of $\tilde{X}$, which was written as P_A and assumed given throughout the earlier text, is the appropriate submatrix from equation (50), and is readily given as

$$P_A = E[\tilde{X}\tilde{X}^T] = \left[H_X Q_A^{-1} H_X^T - H_X Q_A^{-1} H_Z^T \left[H_Z Q_A^{-1} H_Z^T \right]^{-1} H_Z Q_A^{-1} H_X^T \right]^{-1} \tag{52}$$

When all the initial measurements are functions of X only, then

$$P_A = \left[H_X Q_A^{-1} H_X^T \right]^{-1}$$

In this case, it follows by using equations (40) that the matrix, $X_u^T P_A^{-1} X_u$, of equation (38c) is diagonal and yields the calibration accuracy limits:

$$\mu_{ll} = \left(\sum \frac{l_i^2(X_O)}{q_i} \right)^{-1} , \quad \mu_{mm} = \left(\sum \frac{m_i^2(X_O)}{q_i} \right)^{-1} , \quad \mu_{\tau\tau} = \left(\sum \frac{\tau_i^2(X_O)}{q_i} \right)^{-1} \tag{53}$$

The summations in equation (53) are taken over the appropriate category of initial calibrations and $\{q_i\}$ are the MS measurement noise for these

measurements. The meter sticks for this case are given by evaluating equation (44). Thus, for length,

$$\nabla l^* = l^*(X_o) \left(\sum \frac{\hat{l}_i^2}{q_i} \right)^{-1} \sum \frac{\hat{l}_i}{q_i} \nabla l_i$$

and it can be verified that

$$l^*(X) = L_o \sum \frac{\hat{l}_i}{q_i} l_i(X)$$

satisfies the gradient equation, with L_o an arbitrary length factor. Any length proportional to the weighted combination, $\sum (\hat{l}_i/q_i) l_i(X)$, is estimated initially to the accuracy, μ_{ll}. Analogous results are obtained for $m^*(X)$ and $\tau^*(X)$.

The quantities, μ_{ll}, μ_{mm}, $\mu_{\tau\tau}$, in equations (53) depend only on the accuracies of the initial calibrations. They are smaller in magnitude than the MS error ratio of any single calibration if more than one calibration is made. If only one quantity is calibrated initially for a particular unit, then it is the unobservable meter stick for that unit. If more than one is calibrated, then none of them is the meter stick, although the most accurately calibrated one may dominate the appropriate sum in equations (53) and be the approximate meter stick for its unit.

In the preceding special case a direct relation between calibration accuracy limits and initial calibrations is obtained. However, the condition that the initial measurements be functions of X only is not generally sat- isfied and the quantities actually calibrated may have little to do with the variables of interest, X. For example, in orbit determination problems the actual distance calibrations are buried in the multitude of experiments from which the value of the speed of light is estimated. This value is, in turn, used in the initial estimates of such things as radar station location and frequency, gravity field constants. In such cases all the initial measure- ments can influence the accuracy limits, μ_{ll}, μ_{mm}, $\mu_{\tau\tau}$. These limits are the diagonal elements of $\left[X_u^T P_A^{-1} X_u \right]^{-1}$ which, for the general case, can be written as

$$\left[X_u^T P_A^{-1} X_u \right]^{-1} = \left[D - R^T I_Z^{-1} R \right]^{-1} \tag{54}$$

where X_u are the calibration states for X (eq. (38a)), and D is the diagonal matrix

$$D = \begin{bmatrix} \sum l_i^2(Y_o)/q_i & & \\ & \sum m_i^2(Y_o)/q_i & \\ & & \sum \tau_i^2(Y_o)/q_i \end{bmatrix}$$

Also, I_z and R are

$$I_z = H_z Q_A^{-1} H_z^T$$

$$R = \left[H_{z\ell} Q_\ell^{-1} \bar{\ell} \mid H_{zm} Q_m^{-1} \bar{m} \mid H_{z\tau} Q_\tau^{-1} \bar{\tau} \right]$$

The matrix R has one column for each unit other than time; $H_{z\ell}$, H_{zm}, $H_{z\tau}$ are the partitions of H_z in equation (51); Q_ℓ, Q_m, Q_τ are the corresponding partitions of the noise matrix, and $\bar{\ell}$, $\bar{m}$, $\bar{\tau}$ are the vectors

$$\bar{\ell} \equiv \begin{pmatrix} \ell_1(Y_o) \\ \ell_2(Y_o) \\ . \\ . \\ . \end{pmatrix}, \quad \bar{m} \equiv \begin{pmatrix} m_1(Y_o) \\ m_2(Y_o) \\ . \\ . \\ . \end{pmatrix}, \quad \bar{\tau} \equiv \begin{pmatrix} \tau_1(Y_o) \\ \tau_2(Y_o) \\ . \\ . \\ . \end{pmatrix}$$

The elements of the matrix on the left in equation (54) appear in the earlier results (eqs. (42), (44)) as $\mu_{\ell\ell}$, μ_{mm}, $\mu_{\ell m}$, ... All the initial measurements can appear in these elements because of the term, $R^T I_z^{-1} R$.

Finally, it can be shown from equation (54) that the calibration accuracy limits satisfy

$$\mu_{\ell\ell} \geqslant \left(\sum \frac{\ell_i^2(Y_o)}{q_i} \right)^{-1} \tag{55}$$

and similarly for μ_{mm}, $\mu_{\tau\tau}$. The equal sign is removed whenever R has no trivial column; that is, whenever any initial calibration has nontrivial dependence on the variables of Z.

Appendix D contains a simple example to illustrate the use of the above results.

<center>Clock Calibration, the Velocity Standard, and
Estimation of the Speed of Light</center>

The time calibration state was ignored in the earlier discussion. The internationally defined second is based on transitions between energy levels in Cesium-133 (ref. 14). If a cesium clock is used, all time measurements are calibrations and an unobservable time does not occur in the problem.

Assume, then, that a clock different from a cesium clock is in use and that when the clock indicates the passage of one clock second, r seconds on a cesium clock (standard time) have actually passed. Also assume that

$$\hat{r} = 1 \text{ second} \quad \text{and} \quad \dot{r} = 0 \tag{56}$$

In general the clock rate will drift ($\dot{r} \neq 0$) but it is assumed here that the drift is negligible over the time interval of the observations.

The clock rate, r, is analogous to the length of a meter stick in that a reading of the clock time, τ, is a dimensionless comparison

$$\tau = \frac{t}{r} \tag{57}$$

of the standard time, t, with the clock rate, r, in the same way that length measurements using a meter stick of length, l^*, is a dimensionless comparison of two lengths

$$m = \frac{l}{l^*}$$

The state equation

$$\frac{dX}{dt} = \bar{f}(X) \tag{6a}$$

governs the evolution of the state as seen in standard time. As seen in clock time, the state satisfies

$$\frac{dX}{d\tau} = r\bar{f}(X)$$

the state can now be augmented with the unknown clock rate

$$\frac{dy}{d\tau} = \frac{d}{d\tau}\begin{pmatrix} x \\ r \end{pmatrix} = \begin{pmatrix} r\bar{f}(X) \\ 0 \end{pmatrix} = \bar{g}(y)$$

The time calibration state- As before, assume that the time unit is scaled arbitrarily

$$\sec = \lambda \sec'$$

Then all unknowns, as seen in the new unit, are scaled accordingly

$$v' = \lambda^{-1}v \quad \text{frequencies and velocities}$$

$$r' = \lambda r \quad \text{clock rate}$$

$$\tau' = \tau \quad \text{time-dimensionless quantities}$$

43

whence

$$Y' = \Lambda_T Y$$

$$\Lambda_T = \left[\begin{smallmatrix} \lambda^{J_i} \end{smallmatrix} \right]$$

$\{J_i\}$ = powers of the time unit in variables of Y

As before, it is readily shown (noting that the independent variable, τ, is unaffected by the scale change) that

$$\Lambda_T \bar{g}(Y) = \bar{g}(\Lambda_T Y)$$

Assume, also, that all measurements to be processed are time dimensionless, so that

$$m_i(\Lambda_T Y) = m_i(Y) \qquad i = 1, 2, \ldots \qquad (58)$$

Then, using statement [12], page 29, the following state

$$\Sigma_O \equiv J_T Y_O \qquad (59)$$

where

$$J_T = \left[\begin{smallmatrix} J_i \end{smallmatrix} \right]$$

$$Y_O = \begin{pmatrix} X_O(t_O) \\ 1 \end{pmatrix}$$

$X_O(t_O)$ = reference initial state

is an unobservable state (the time calibration state) of the linearized augmented system.

The linearized augmented state equation- This equation can be routinely derived as

$$\frac{dy}{d\tau} = G(Y_O)y$$

where

44

$$G(Y_o) = \begin{pmatrix} \vdots \\ \nabla^T g_i(Y) \\ \vdots \\ \vdots \end{pmatrix}_{Y_o(\tau)} = \left[\begin{array}{c|c} \hat{r}F(X_o(\tau)) & \dot{X}_o(\tau) \\ \hline 0 \ . \ . \ . \ 0 & 0 \end{array} \right] \tag{60}$$

Here, $F(X_o)$ is the coefficient matrix of the linearized equation in standard time.

An alternate derivation can be made by defining the state deviation

$$\xi(\tau) = X\big(t(\tau)\big) - X_o(\tau) = X(\tau) - X_o(\tau) + \dot{X}_o(\tau)\big(t(\tau) - \tau\big) + \ . \ . \ .$$

or

$$\xi(\tau) = x(\tau) + \dot{X}_o(\tau)\tau\delta r + \ . \ . \ . \tag{61}$$

where $x(\tau)$ is the usual definition of the state deviation $(X(\tau) - X_o(\tau))$, which, to first order, is a solution of

$$\frac{dx}{d\tau} = F\big(X_o(\tau)\big)x$$

Equation (60) is now readily reproduced by taking the τ-derivative of equation (61). Thus, the augmented linearized equation describes the time history of $\xi(\tau)$ (eq. (61)) and includes the effect of the secularly increasing discrepancy between standard time and clock time.

Linearized measurements- If a data type, $M(Y)$, is sampled the linearized data used in the linearized estimation procedure can be routinely expressed as

$$\delta m = (\nabla M)^T_{Y_o(\tau)} y(\tau) + \varepsilon$$

or

$$\delta m = h(\tau)^T y(t_o) + \varepsilon$$

where

$$h(\tau) = \phi^T(\tau, t_o)(\nabla M)^T_{Y_o(\tau)}$$

An alternate derivation can be given as follows. Suppose a data type, $M(X)$, is sampled at the clock time, τ. The sampling obtained is then

$$m = M\Big(X\big(t(\tau)\big)\Big) + \varepsilon$$

That is, it is a (noisy) sampling of $M(X)$ on the actual state at the standard time corresponding to the clock time, τ. The residual or linearized sampling is then computed by subtracting the value of $M(X)$ evaluated on the reference state at the time $t = \tau$.

$$\delta m = M\Big(X\big(t(\tau)\big)\Big) - M\big(X_0(\tau)\big) + \varepsilon$$

which, after expanding $X\big(t(\tau)\big)$ and retaining first-order terms, is

$$\delta m = \nabla M^T \xi(\tau) + \varepsilon$$

where $\xi(\tau)$ is the state deviation defined in (61). It is not necessary to include the clock reading as an additional measurement in the problem, since any random reading error, ε_r, can be included with the sampling error, ε, in the amount, $\nabla M^T \dot{X}_0(\tau) \varepsilon_r$.

Velocity calibration accuracy- If all data types are dimensionless in distance and time, an unobservable velocity exists which serves as the standard against which all velocities in the problem are compared. Its initial estimation accuracy is the lower bound on the accuracy to which any velocity can be estimated from any sampling of the data types.

This estimation accuracy limit is given as

$$\eta_v = \mu_{ll} - 2\mu_{l\tau} + \mu_{\tau\tau} \tag{62}$$

where μ_{ll}, $\mu_{l\tau}$, $\mu_{\tau\tau}$ are the elements of the matrix

$$\begin{pmatrix} \mu_{ll} & \mu_{l\tau} \\ \mu_{l\tau} & \mu_{\tau\tau} \end{pmatrix} = \Big[Y_u^T P_A^{-1} Y_u \Big]^{-1}$$

$$Y_u = \Big[\Sigma_l \Sigma_\tau \Big]$$

$$\Sigma_l = J_l Y_0 \qquad \text{distance calibration state}$$

$$\Sigma_\tau = J_\tau Y_0 \qquad \text{time calibration state}$$

$$P_A \qquad\qquad \text{initial covariance matrix}$$

Equation (62) for the velocity accuracy limit is obtained by applying equation (42c). It can also be shown that the velocity standard in the problem, v^*, is proportional to l^*/t^* where l^*, t^* are the distance and time meter sticks of the problem.

Speed of light estimation in space mission problems- Space probe trajectory estimation problems have been treated as experiments to estimate a number of fundamental parameters of the solar system. One expects that a necessary condition for significant improvements in the estimates of such parameters is that their initial estimation accuracies be significantly poorer than the appropriate calibration accuracies in the experiment.

For example, consider the possibility of improving the estimate of the speed of light in a space probe trajectory estimation problem using radar measurements. The independent unknown quantities required to define the data types can be listed, in general, as

$$
Y = \begin{pmatrix} c/\hat{c} \\ r \\ \overline{R}_v \\ \overline{V}_v \\ \overline{R}_{si} \\ f_{si} \\ \cdot \\ \cdot \\ \cdot \\ w \\ \mu \\ J_i \\ \cdot \\ \cdot \\ \cdot \\ \overline{R}_{P_i} \\ \overline{V}_{P_i} \\ \mu_i \\ J_{ij} \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}
\begin{array}{l}
\text{Speed of light} \\[4pt]
\text{Clock rate} \\[4pt]
\text{Vehicle state} \\[14pt]
\text{Radar station locations and frequencies} \\[40pt]
\text{Earth rotation rate and gravity field parameters} \\[60pt]
\text{Planet states and gravity field} \\ \quad\text{parameters}
\end{array}
$$

The initial measurements include an experiment to estimate the speed of light

$$m_1 = d = c\Delta t$$

$$m_2 = \Delta t$$

This experiment consists of measuring a distance, d, and the time, Δt, required for light to travel this distance. Both measurements are calibrations and the experiment $\{m_1, m_2\}$ is equivalent to a large number of similar experiments from which the current estimate of the speed of light is obtained. The initial estimation accuracy for the speed of light is then

$$q_c = E\left[\left(\frac{\tilde{c}}{\hat{c}}\right)^2\right] = \frac{q_1}{\hat{d}^2} + \frac{q_2}{\Delta\hat{t}^2}$$

where q_1, q_2, are noise variances of the two calibrations. Currently, the speed of light is estimated to one part in 300,000 or $\sqrt{(q_c)} = 0.3\times10^{-6}$.

All remaining initial measurements are assumed to be functions of Y and these may include additional distance and time calibrations

$$\ell_{i_\ell}(Y) \qquad i_\ell = 1, \ldots, N_\ell$$

$$t_{i_\tau}(Y) \qquad i_\tau = 1, \ldots, N_\tau$$

The complete set of initial measurements depends on Δt as well as the variables of Y. In that case, temporarily augment the state to

$$Y' = \begin{pmatrix} \Delta t \\ \hline Y \end{pmatrix}$$

and then the initial covariance of Y' is given as

$$P_A(\tilde{Y}') = \left[H_A Q_A^{-1} H_A^T\right]^{-1}$$

$$H_A = \begin{bmatrix} \hat{c} & 1 & 0 & \ldots & 0 \\ \hline \Delta t\hat{c} & 0 & & & \\ 0 & 0 & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ 0 & 0 & & & \end{bmatrix} \quad \text{and} \quad Q_A = \begin{bmatrix} q_1 & & & \\ & q_2 & & \\ & & \cdot & \\ & & & \cdot \end{bmatrix}$$

It is assumed that the rank of H_A is equal to the number of variables in Y'.

The variable, Δt, is irrelevant to the estimation problem since the data types are independent of it (they are functions of Y). Neither is it of interest to estimate Δt via the correlations of its initial estimation error with those of the remaining unknowns. In that case, Δt, can be removed from the estimation equations as described earlier and, in particular, equation (54) is applied to obtain the calibration accuracies of the reduced problem. The result for the velocity accuracy limit is (lengthy algebra is omitted here)

$$n_V = \frac{q_c}{1 + \frac{q_c}{\sigma_V}} = \left(\frac{1}{q_c} + \frac{1}{\sigma_V} \right)^{-1} \tag{63}$$

where

$$\sigma_V \equiv \left(\sum_{i=1}^{N_\ell} \frac{\hat{\ell}_i^2}{q_i} \right)^{-1} + \left(\sum_{i=1}^{N_\tau} \frac{\hat{t}_i^2}{q_i} \right)^{-1}$$

The summations in σ_V are taken, respectively, over all initial length and time calibrations except those in the speed of light experiment $\{m_1, m_2\}$. It is apparent from the result that

1. If the speed of light experiment had not occurred then the velocity accuracy limit would be σ_V. Conversely, if no calibrations other than $\{m_1, m_2\}$ had occurred, then the velocity accuracy limit would be q_c and the speed of light would be the unobservable velocity standard.

2. In general, $n_V < q_c$ and it is theoretically possible to improve the accuracy of the speed of light estimate to n_V by sampling noncalibration data types of the form $m(Y)$.

3. A necessary condition for significant improvement in the speed of light estimate by sampling noncalibration data types, $m(Y)$, is, roughly, that

$$\frac{q_c}{\sigma_V} > 1$$

that is, that the velocity calibration accuracy due to the remaining calibrations, $\{\ell_i\}$, $\{t_i\}$ be at least as good as the accuracy of the initial speed of light estimate, q_c.

4. In practice, it is usually true that

$$\frac{q_c}{\sigma_V} \ll 1$$

so that negligible improvement can be expected and the speed of light is very nearly unobservable in space vehicle trajectory estimation problems.

Discussion

An elementary view of estimation is taken here to analyze the naturally occurring calibration limit on the obtainable estimation accuracy. Estimation, in general, is a comparison of some set of unknown quantities with the internationally defined unit quantities: the second, meter, kilogram, and degree Celsius. A direct comparison of some quantities in the problem with the unit quantities (calibrations) must occur at some point in the estimation procedure. In most cases this is done initially and subsequent measurements are comparisons among unknown quantities. In that case, the accuracy to which the unknowns can be estimated in terms of the international units is limited by the accuracy of the initial calibrations. For example, the accuracy with which a length can be estimated by comparing it to a meter stick is limited by the accuracy with which the meter stick was calibrated initially.

Estimation problems are generally analogous to the above simple example even though several data types or measuring devices, many initial calibrations, and other physical units besides the unit of length may appear in the problem. If the data types are not calibrations, then an independent unobservable state is readily given, one for each physical unit in the problem (statements [12] and [13] and eq. (59)). These states define an intermediate standard of comparison, or meter stick (eq. (44)), and a calibration accuracy (eq. (42)) for each physical unit (time, distance, mass, and temperature) having the best initial estimation accuracy in the problem. The procedure of estimating the unknowns by sampling noncalibration data types is, at most, one of comparing all times, lengths, masses, and temperatures in the problem to these meter sticks.

The initial covariance and the calibration accuracies can be given in terms of the *initial measurements* (eqs. (49) to (52)). In the simplest case, the accuracy limits depend only on the initial calibration measurements (eq. (53)) but, in general, these limits can depend on all initial measurements (eq. (54)). An example is given in the discussion of the velocity accuracy limit in space probe trajectory estimation (eq. (63)).

Unobservable calibration states are theoretically present whenever the measurements are not calibrations, but these states are often eliminated in practice by treating some quantities, which have been estimated very accurately initially as exactly known quantities. For the theory of this report, the values of all physical quantities in a problem are considered unknowns; that is, all are estimated initially but their true values are unknown and their estimation errors can affect the data. The state variables to be estimated are then any set of independent unknowns which suffice to define all unknown quantities in the problem. In practice, variables which are very accurately estimated initially are often deleted from the state variables; that is, they are treated as having no estimation error. It is assumed that the inclusion of such variables in the estimation would have negligible effect on the estimates of the remaining state variables. As a result of this

50

procedure, the sufficient conditions for the existence of unobservable calibration states (statement [12]) are no longer satisfied by the problem; the required homogeneity properties of the state equation are eliminated and the data types can become calibrations or comparisons with exactly known quantities. The resulting removal of the lower bounds on estimation accuracy causes no difficulty in practice provided reported accuracies do not approach the actual calibration accuracy limit. It is, of course, not possible to exceed this limit without performing new calibrations. The size of the calibration accuracy limit is given from the initial covariance (eq. (42)) and can be estimated if some reasonably correct idea of the initial measurements is available.

Generally, a necessary condition for significant improvement in the estimate of any parameter is that its initial estimation accuracy be significantly poorer than the appropriate calibration accuracy limit. At best, the parameter can be estimated to nearly the calibration accuracy limit by indefinite sampling of the data types. However, the existence of unobservable states in addition to the calibration states can impose further limits on the accuracy to which a parameter can be estimated.

RÉSUMÉ

Mathematically, estimation is a matter of processing noisy measurements to obtain the best estimate of the unknown quantities on which the measurements depend. The elements which appear in the processing equation for linear estimation (eqs. (5)) are the prior estimate and covariance and the measurements (costates, MS noise, and data). The processing results in a new estimate of the state and its covariance of estimation errors. This report describes some results from a study of the relations between the elements of the processing equation and the resulting estimate and performance.

In the first section, the processing of arbitrary sequences of measurements is examined using the notion of equivalent sequences and data. This notion provides a basis for carrying out a number of operations in the processing of measurements. A sequence can be compressed to an equivalent sequence containing a minimum number of measurements. The uncorrelated equivalent basis sequence and its associated eigenvalues give the performance of the sequence in reducing the estimation errors from their values prior to processing. Further, equivalent sequences can be used to selectively reduce the information to be processed so as to estimate only some of the state variables, or to separate the processing equations into independent lower order parts.

In many estimation problems, the measurements are obtained by sampling one or more data types at an arbitrary sequence of times. The characteristics of the data types are then constraints on any such sampling. One characteristic is observability, the sufficiency of the data types for the determination of the state variables. If the data types are insufficient, then a suitable transformation of the state gives the data types in terms of a

reduced set of new variables which are linear combinations of the original variables and which can be determined from some sampling of the data type. This transformation also separates the estimation calculations into trivial and nontrivial parts, the latter being a lower order set of equations for estimating the reduced set of observable unknowns.

A theory of calibration accuracy limits in estimation problems is obtained by applying observability analysis and recognizing the physical nature of estimation. Estimation is intended to compare a physical system with the internationally defined physical units and this is done by taking measurements which, in general, are comparisons of like physical phenomena. If the measurements are not calibrations (not direct comparisons with the unit physical phenomena), then the estimation procedure is shown to be one of comparing the system to a set of intermediate standards of time, length, mass, or temperature. These intermediate standards are unobservable and their initial estimation accuracies (calibration accuracies) limit the accuracy to which any time, length, mass, or temperature in the problem can be estimated without performing new calibrations. Further, to obtain significant improvement in the estimate of any unknown, it is necessary that its initial estimation accuracy be significantly poorer than the appropriate calibration accuracy. The existence of independent unobservable parameters in addition to the intermediate standards will further limit the obtainable estimation accuracy.

In principle the state is either observable to the data types or the estimation problem can be reduced to a lower order problem in which a reduced number of variables are observable. Computations to determine the observable parameter space and carry out the reduction are based on the information integral (eq. (19)). An application of this method is found in reference 12. The necessary computations are difficult, especially in large-order problems. Alternative methods based on derivatives of the data type (ref. 13) are available but not yet explored. The structure of relations between data and performance in the case that all states are observable remains unclarified. In many orbit determination problems using radar data it is recognized that one or more parameters may be observable but computationally difficult to estimate (e.g., refs. 4, 5, 12) because the data type provides little information on these parameters compared to others. Computationally, such cases can be treated with special procedures (e.g., refs. 5, 12).

Ames Research Center
      National Aeronautics and Space Administration
            Moffett Field, Calif., 94035, January 29, 1971

APPENDIX A

COMPUTATION OF AN EQUIVALENT BASIS SEQUENCE

Given an information matrix

$$I_f = HQ^{-1}H^T$$

obtain an equivalent basis; that is, a matrix, of maximal rank such that

$$H_E H_E^T = I_f$$

This problem is identical to the computation of a square root matrix of maximal rank for any symmetric positive semidefinite matrix. The method reviewed here is also described in reference 10.

COMPUTATIONAL METHOD

Let B be any $n \times n$ symmetric positive semidefinite matrix and let $\{\bar{v}_k\}$ be the columns of any nonsingular $n \times n$ matrix. Define the matrices $\{B_k\}$ as

$$B_1 = B, \qquad B_{k+1} = \begin{cases} B_k - \dfrac{B_k \bar{v}_k \bar{v}_k^T B_k}{\bar{v}_k^T B_k \bar{v}_k} & \text{if } \bar{v}_k^T B_k \bar{v}_k \neq 0 \\[4mm] B_k & \text{if } \bar{v}_k^T B_k \bar{v}_k = 0 \end{cases} \tag{A1}$$

and define the vectors $\{\bar{c}_k\}$ as

$$\bar{c}_k = \begin{cases} \dfrac{B_k \bar{v}_k}{\sqrt{\bar{v}_k^T B_k \bar{v}_k}} & \text{if } \bar{v}_k^T B_k \bar{v}_k \neq 0 \\[4mm] 0 & \text{if } \bar{v}_k^T B_k \bar{v}_k = 0 \end{cases} \tag{A2}$$

The matrix whose columns are the nontrivial vectors among $\{\bar{c}_k, k = 1, \ldots, n\}$ is a square root of B having maximal rank.

53

This result is proved in the following series of statements.

[1] The matrices B_1, B_2, . . . are all symmetric positive semidefinite. This follows from induction since the statement is true for B_1 and if it is assumed true for B_k it follows that it is true for B_{k+1}.

[2] The matrix B_{n+1} is zero,

$$B_{n+1} = 0$$

The null space of B_{k+1} includes (a) the vector, $\bar{v}_k$ and (b) the null space of B_k. This can be verified from the definition (A1) above; it follows immediately that $B_{k+1}\bar{v}_k = 0$ and that if $\bar{v}$ is any vector for which $B_k\bar{v} = 0$ then also $B_{k+1}\bar{v} = 0$. Therefore, the null space of B_{k+1} includes the vectors $\bar{v}_1$, $\bar{v}_2$, . . ., $\bar{v}_k$ and, in particular, the null space of B_{n+1} includes all n independent vectors, $\{\bar{v}_k\}$, and is zero.

[3] Let C be the matrix whose columns are the nontrivial vectors among $\{\bar{c}_k, k = 1, . . ., n\}$. C is a square root matrix of B

$$B = CC^T \tag{A3}$$

Employing statement [2] and the definition of B_k, $\bar{c}_k$, and C obtain

$$0 = B_{n+1} = B_n - \bar{C}_n\bar{c}_n^T = . . . = B_1 - \sum_{k=1}^{n} \bar{c}_k\bar{c}_k^T = B - CC^T$$

[4] The nontrivial vectors among $\{\bar{c}_k, k = 1, . . ., n\}$ are independent and C has maximal rank.

To prove this, show that there is no nontrivial linear combination of the nontrivial $\{\bar{c}_k\}$ which is zero. Let

$$\sum_{k=1}^{n} s_k\bar{c}_k = 0 \tag{A4}$$

be any linear combination which is zero and consider the dot product of $\bar{v}_1$, with this sum of vectors

$$\left(\sum_{k=1}^{n} s_k\bar{c}_k^T\right)\bar{v}_1 = 0 \tag{A5}$$

for nontrivial $\bar{c}_k$ we have from (A2)

$$\bar{v}_1^T \bar{c}_k = \frac{\bar{v}_1^T B_k \bar{v}_k}{\left(\bar{v}_k^T B_k \bar{v}_k\right)^{1/2}}$$

But, as pointed out in statement [2], the vector $\bar{v}_1$ is in the null space of all $\{B_k, k \geq 2\}$ and all B_k are symmetric positive semidefinite so that

$$\bar{v}_1^T B_k = 0 , \qquad k \geq 2 \qquad \text{(A6)}$$

and equation (A5) becomes

$$s_1 \bar{v}_1^T \bar{c}_1 = 0$$

Thus, if $\bar{c}_1$ is nontrivial, then s_1 is zero and equation (A4) becomes

$$\sum_{k=2}^{n} s_k \bar{c}_k = 0 \qquad \text{(A7)}$$

These steps are repeated for $\bar{v}_2$, forming the dot product of $\bar{v}_2$ with equation (A7) and obtaining $s_2 = 0$ if $\bar{c}_2$ is nontrivial. When repeated in succession for $\bar{v}_3$, $\bar{v}_4$, . . ., $\bar{v}_n$, the $\{s_j\}$ are zero for all nontrivial $\bar{c}_j$. Consequently, there is no nontrivial linear combination of the nontrivial $\{\bar{c}_k\}$ which is zero. These vectors are, therefore, independent and the matrix C has maximal rank.

In summary, the matrix C whose columns are the nontrivial $\{\bar{c}_k\}$ is an equivalent basis of any sequence whose information matrix is B.

CHOICE OF THE VECTORS $\{\bar{v}_k\}$

The vectors $\{\bar{v}_k\}$ can be any n independent vectors. The columns of the unit matrix are utilized in reference 10 and this choice places the resulting square root in lower triangular form.

Alternatively, the $\{\bar{v}_k\}$ can be selected to remove beforehand those which yield trivial vectors in equation (A2). This can be done by obtaining a partitioned matrix

$$V = \left[V_u \mid V_m \right]$$

in which the columns of V_u are any basis of the solution space of

$$B\bar{v} = 0 \qquad \text{(A8)}$$

and the columns of V_m are any completion to a set of n independent vectors, for example, a basis of the solution space of

$$V_u^T \bar{v} = 0$$

It is readily seen that the number of columns of V_m is the rank of B, and that all columns of V_u give trivial vectors, $\bar{C}_k$, when processed in equations (A1) and (A2). In that case, only the columns of V_m need be processed to obtain the square root of B and these all yield nontrivial vectors in equation (A2).

Insensitivity to computational error is an important factor in the choice of computational procedure for this problem, but this question is not addressed here.

INFORMATION SEPARATION

## SEPARATION LEMMA

Let (H, Q) be a measurement sequence and $\mathcal{H}_m$ the column space of H. Let $\mathcal{H}_A$ be any arbitrary subspace of $\mathcal{H}$ and

$$\dim(\mathcal{H}_A \cap \mathcal{H}_m) = k$$

Then an equivalent basis sequence of (H, Q) exists which has k elements that are a basis of $\mathcal{H}_A \cap \mathcal{H}_m$.

The equivalent bases of (H, Q) are all the square roots of maximal rank of its information matrix. All such square roots are bases of the same space, $\mathcal{H}_m$. The statement to be proved asserts that for any specified subspace of $\mathcal{H}_m$, a square root can be found, part of which is a basis of the specified subspace. Such a separated equivalent basis is not unique for the specified subspace, but it separates the information matrix uniquely.

$\mathcal{H}_A \cap \mathcal{H}_m = \mathcal{H}_{Am}$
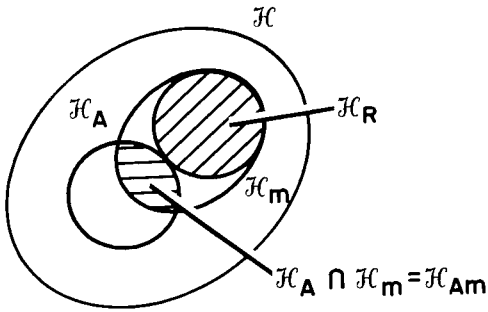
The sketch visualizes the lemma. Here, $\mathcal{H}$ is the whole costate space and includes both $\mathcal{H}_A$ and $\mathcal{H}_m$. A partitioned equivalent primary basis

$$(H, Q) \doteq \left[H_{AM} \mid H_R \right]$$

is to be found for which H_{AM} is a basis of $\mathcal{H}_A \cap \mathcal{H}_m$ and H_R refers to the remaining columns which are a basis of some subspace of $\mathcal{H}_m$ outside $\mathcal{H}_A$.

PROOF

The following proof outlines the required computational procedure.

1. Let H_A, H_m be, respectively, any basis of $\mathcal{H}_A$ and an equivalent primary basis of (H, Q). The first is assumed given as the definition of $\mathcal{H}_A$ and the second is computed by procedures already given and always exists. Assume, also, that H_m has r columns.

2. Compute any basis, H_{AM_1}, of the subspace $\mathcal{H}_{AM}$. The intersection, $\mathcal{H}_{AM}$, is always a linear vector space so that a basis exists and can be

computed as follows. Let $\left[H_A \mid H_B\right]$ be any completed basis of the whole space, $\mathcal{H}$, which contains H_A, and express H_m in terms of this basis

$$H_m = \left[H_A \mid H_B\right] \left[\begin{array}{c} A \\ \hline B \end{array}\right] \tag{B1}$$

where the coefficient matrix, $\left[\begin{array}{c} A \\ \hline B \end{array}\right]$, is computed by inverting (B1). Then define

$$D_1 \equiv \text{any basis of } \{\bar{d}: B\bar{d} = 0\} \tag{B2}$$

and then

$$H_{AM_1} \equiv H_m D_1 = H_A A D_1 \tag{B3}$$

Every column of H_{AM_1} is simultaneously in $\mathcal{H}_A$ and $\mathcal{H}_m$; every column is independent, and every vector which is both in $\mathcal{H}_A$ and $\mathcal{H}_M$ can be given in terms of H_{AM_1}. Consequently, H_{AM_1} is a basis of $\mathcal{H}_{AM}$.

3. Next, define the partitioned orthogonal $r \times r$ matrix

$$[D \mid S]$$

where D is obtained as the Gram-Schmidt orthogonalization of the columns of D_1, and S is any orthonormal basis of $\{\bar{s}: D^T \bar{s} = 0\}$. Then define

$$H_m^* = H_m [D \mid S] \tag{B4}$$

Since $[D \mid S]$ is orthogonal then

$$H_m^* \doteq H_m \tag{B5}$$

and the submatrix

$$H_{AM} \equiv H_m D$$

is a basis of $\mathcal{H}_{AM}$, while

$$H_R \equiv H_m S \tag{B6}$$

is such that $\mathcal{H}_R \cap \mathcal{H}_A = 0$. Consequently,

58

$$H_m^* = \left[H_{AM} \mid H_R \right] \tag{B7}$$

is the required partitioned equivalent basis of (H, Q). The corresponding equivalent data for the fictitious measurements defined by the columns of (B7) is given from equation (12) of the text.

DISCUSSION

1. The separation procedure is trivial in the case of sequential processing of single measurements. A single measurement is either in or outside of $\mathcal{H}_A$, so that one part of the separated basis is the measurement itself and the other part is null.

2. Given the sequence, (H, Q), and the subspace, $\mathcal{H}_A$, the information matrix is uniquely separated by the above procedure

$$HQ^{-1}H^T = H_{AM}H_{AM}^T + H_R H_R^T \tag{B8}$$

The corresponding parts of any two partitioned equivalent bases are also equivalent since they are constructed to span the same subspaces (statement [6] of the text).

3. The separation states that the equivalent basis can be selected so that some of its measurements are in any specified subspace and the remainder are outside that subspace. This permits selective processing of information. The available information is fixed by the actual sequence of measurements but this information can be reduced by deleting or delaying the processing of any measurement(s) from the actual sequence or from any equivalent sequence. Finally, the separation permits removal of measurements within or outside any selected subspace by deleting from processing either of the terms in (B8).

4. The influence of each of the separated terms in (B8) on performance (reduction of MS estimation errors from their prior values) is considered next. Suppose $\mathcal{H}_A$ is some given subspace of $\mathcal{H}$ and let $\mathcal{H}_B$ be its uncorrelated complement space

$$\mathcal{H}_B \equiv \{h: \quad h^T P_o H_A = 0\} \tag{B9}$$

P_o being the prior covariance. If only the term, H_{AM}, is processed then there will be nontrivial reduction of MS errors for every parameter whose costate is not in the uncorrelated complement space of $\mathcal{H}_{AM}$. This includes $\mathcal{H}_{AM}$ and may or may not include all of $\mathcal{H}_A$. There is no error reduction for any parameter whose costate is in the complement space of $\mathcal{H}_{AM}$ and this includes $\mathcal{H}_B$.

59

If only the measurements H_R are processed, then there is nontrivial error reduction for some parameters with costates in $\mathcal{K}_B$ since every column of H_R is outside $\mathcal{K}_A$. In addition, there is nontrivial error reduction for some parameters with costates in $\mathcal{K}_A$ except in the special case that every column of H_R is in $\mathcal{K}_B$. Thus, the capacity of the original sequence to reduce errors in $\mathcal{K}_A$ is not generally isolated in the extracted measurements, H_{AM}.

As noted above, if only the extracted measurements, H_{AM}, are processed then there is no error reduction for parameters with costates in $\mathcal{K}_B$. Further, the posterior estimation errors for parameters with costates in $\mathcal{K}_A$ are mutually uncorrelated with those in $\mathcal{K}_B$ so that the complement space of $\mathcal{K}_A$ is the same set of parameter costates for both the prior and posterior covariances. Consequently, the complement space of $\mathcal{K}_A$ remains fixed if a succession of measurement sequences are treated in the same way. It is possible, therefore, to prevent error reduction on any given subspace, $\mathcal{K}_B$, by selectively processing only information in its complement space, $\mathcal{K}_A$, and the complement space need be determined only once.

SEPARATION OF ESTIMATION PROBLEMS INTO INDEPENDENT PARTS

The estimation calculations can be forced to separate into independent parts by selective removal of some information from each sequence that is processed. The data processing then separates into two lower order problems.

Let P_A be the initial covariance (MS estimation errors of the reference trajectory), let $\mathcal{K}_A$ be any given set of parameters with basis, H_A, whence its initial uncorrelated complement space is

$$\mathcal{K}_B \equiv \{h: \ h^T P_A H_A = 0\} \tag{B10}$$

The two subspaces, $\mathcal{K}_A$, $\mathcal{K}_B$ form a complete decomposition of the whole costate space and will remain fixed throughout the data processing.

Suppose that the following two-part separation is made on every measurement sequence (H, Q) to be processed

$$(H, \ Q) \doteq \left[H_{AM} \ \vdots \ H_{RB} \ \vdots \ H_R' \right] \tag{B11}$$

where $\left[H_{AM} \ \vdots \ H_R \right]$ is the first separation of (H, Q) into a basis of $\mathcal{K}_{AM}$ and a remainder H_R; and $\left[H_{RB} \ \vdots \ H_R' \right]$ is a separation of H_R into a basis of $\mathcal{K}_R \cap \mathcal{K}_B$ with H_R' as the remainder from this second separation. If H_R' is removed from every sequence to be processed then $\mathcal{K}_A$ is unobservable to the measurements H_{RB}, and $\mathcal{K}_B$ is unobservable to H_{AM}, so that the estimation

60

calculations can be decomposed into two lower order problems associated with each of the two subspaces. Some formulas for the separated problem are given next. Algebra is largely straightforward and omitted.

Let H_A, H_B be any bases of $\mathcal{H}_A$, $\mathcal{H}_B$, respectively, and define the corresponding subspaces of the state space with bases

$$\left.\begin{array}{c} X_A = P_A H_A \\ X_B = P_A H_B \end{array}\right\} \tag{B12}$$

Together these form a transformation of the state space

$$x = \left[X_A \mid X_B\right] w = X_A w_A + X_B w_B \tag{B13}$$

The new variables, w, are the components of the state, x, in the basis $\left[X_A \mid X_B\right]$, and these are linear combinations of the parameters of H_A and H_B, since

$$w_A = \left(H_A^T P_A H_A\right)^{-1} H_A^T x$$

$$w_B = \left(H_B^T P_A H_B\right)^{-1} H_B^T x$$

Transforming the state estimation error, $\tilde{x}$, into the new variables, $\tilde{w}$, obtain the initial covariance of $\tilde{w}$ as

$$W_O = E\left[\tilde{w}\tilde{w}^T\right] = \begin{bmatrix} \left(H_A^T P_A H_A\right)^{-1} & 0 \\ \hline 0 & \left(H_B^T P_A H_B\right)^{-1} \end{bmatrix} \tag{B14}$$

The nontrivial parts of W_O are the initial covariances of $\tilde{w}_A$, $\tilde{w}_B$, and are written W_{OA}, W_{OB} hereafter.

If one measurement sequence is separated as in (B11) and only H_{AM}, H_{RB} are processed then the posterior covariance of $\tilde{w}$ is

$$W = \begin{bmatrix} W_A & 0 \\ \hline 0 & W_B \end{bmatrix} = \begin{bmatrix} \left(W_{OA}^{-1} + M_A M_A^T\right)^{-1} & 0 \\ \hline 0 & \left(W_{OB}^{-1} + M_B M_B^T\right)^{-1} \end{bmatrix} \tag{B15}$$

61

where, for convenience, the notation

$$M_A \equiv X_A^T H_{AM} = H_A^T P_A H_{AM}$$
$$M_B = X_B^T H_{RB} = H_B^T P_A H_{RB}$$

(B16)

has been used. The estimate, $\hat{w}$, is given by

$$\hat{w}_A = W_{OA} M_A \left[I + M_A^T W_{OA} M_A \right]^{-1} y_A$$
$$\hat{w}_B = W_{OB} M_B \left[I + M_B^T W_{OB} M_B \right]^{-1} y_B$$

(B17)

Here, y_A, y_B are the equivalent data vectors corresponding to the measurements H_{AM}, H_{RB}, respectively, and which are assumed computed with the separation of (H, Q) in accordance with equation (12) of the text. If additional measurement sequences are separated as in (B11) and only H_{AM}, H_{RB} are processed from each sequence, then the processing equations remain separated, being the same as (B15) and (B17) except to replace, W_{OA}, W_{OB} with the prior covariances and to replace $\hat{w}_A$, $\hat{w}_B$ with the changes in estimates

$$\hat{w}_A - \hat{w}_A^- , \qquad \hat{w}_B - \hat{w}_B^-$$

Thus, by removing H_R' from each sequence, the calculations for estimating $\tilde{x}$ can be separated into two independent lower order problems for estimating $\tilde{w}_A$, $\tilde{w}_B$. The processing equations (B15) and (B17) are entirely analogous to the original equations (5) in the text.

Finally, the estimate, $\hat{x}$, and covariance, $E[\tilde{x}\tilde{x}^T]$, for the original state variables are recovered at any step from

$$\hat{x} = X_A \hat{w}_A + X_B \hat{w}_B$$
$$P = X_A W_A X_A^T + X_B W_B X_B^T$$

(B18)

The separation above is trivial for sequential processing of single measurements and is intended only for treatment or analysis of one or more batches of measurements. The neglected measurements, H_R', are generally nontrivial and provide nontrivial error reduction for some parameters in both $\mathcal{H}_A$ and $\mathcal{H}_B$, since all the columns of H_R' are neither in $\mathcal{H}_A$ nor $\mathcal{H}_B$ by construction. Further, the information content of all the neglected measurements will differ depending on the subdivision of the total set of measurements into batches for separation. Although the choice of subspace, $\mathcal{H}_A$, can be made

arbitrarily, the importance of the neglected measurements, H_R', to error reduction in $\mathcal{H}_A$, $\mathcal{H}_B$ is expected to depend on how well the choice of $\mathcal{H}_A$ corresponds to some naturally occurring subdivision of the problem into nearly independent parts. Such a natural subdivision, when it exists, would be inherent in the problem; that is, in the observed dynamic system and in the data types employed for measurements, but this matter is beyond the present scope which deals with processing arbitrary measurement sequences.

APPENDIX C

SUFFICIENT CONDITIONS FOR UNOBSERVABLE STATES

The sufficient conditions and corresponding unobservable states derived here are the basis of discussion in the text of the proposition that if data types exclude calibrations, then the physical scale of the problem cannot be determined to better accuracy than it was known *a priori*, and this limits the accuracy to which the state variables can be determined from the data types.

The statement to be proved here is that unobservable states exist if the state equation and data types possess certain homogeneity properties. These properties occur naturally in many problems as follows: the state variables have arbitrary units; compatibility requires that the components of the state equation be homogeneous functions, the consequence of which is that a class of solutions related to the reference trajectory can be given. These solutions can be regarded either as distinct or as the same solution viewed with different scales for the physical units of length, mass, and temperature. If calibrations are excluded the data types are also homogeneous functions which are unchanged by any change in the scale of physical units. Hence identical data are generated by any of the trajectories mentioned above.

The homogeneity properties given below suffice for the existence of trajectories which are indistinguishable to the data types. The corresponding result for the linearized problem is obtained by applying Euler's theorem for homogeneous functions to determine appropriate unobservable states.

HOMOGENEOUS FUNCTIONS

In the following, $f(X)$ is any scalar function of the $n \times 1$ state vector defined on an open set, S, in χ. Homogeneous functions are usually defined as follows (ref. 15, p. 134).

Definition. $f(X)$ is homogeneous of degree p on S provided

$$f(\lambda X) = \lambda^p f(X) \tag{C1}$$

for all x in S and all real λ such that λX is in S.

It is necessary to generalize this definition in order to study homogeneity properties related to scale changes in physical units.

Definition. $f(X)$ is homogeneous of degree p for the integers $\{J_1, \ldots, J_n\}$ provided

$$f(\Lambda X) = \lambda^p f(X)$$

where

$$\Lambda \equiv \left\lceil \lambda^{J_i} \right\rfloor \tag{C2}$$

for all X in S and all real λ such that ΛX is in S.

The integers in this definition can be positive, negative, or zero. The definition (C2) includes (C1) as a special case.

INDISTINGUISHABLE TRAJECTORIES

In the following, $\{m_i(X), i = 1, \ldots, k\}$ are the data types, and the state equation is

$$\dot{X}(t) = \bar{f}(X(t)) \tag{6}$$

The data types and the components of $\bar{f}$ are all assumed defined in an open set, S, in χ and to have continuous partial derivatives.

Two solutions of the state equation $\{X_1(t), X_2(t), t_o \leqslant t \leqslant t_F\}$ are *indistinguishable* provided they generate identical data; that is, provided

$$m_i\big(X_1(t)\big) = m_i\big(X_2(t)\big) \qquad i = 1, \ldots, k; \; t_o \leqslant t \leqslant t_F \tag{C3}$$

Indistinguishable trajectories can arise as follows. Let $\{X_o(t), t_o \leqslant t \leqslant t_F\}$ be any solution of (6).

If there exists a set of integers $\{J_1, \ldots, J_n\}$ such that

$$\left.\begin{array}{ll} (1) & \bar{f}(\Lambda X) = \Lambda \bar{f}(X) \\[1em] (2) & m_i(\Lambda X) = m_i(X); \qquad i = 1, \ldots, k \end{array}\right\} \tag{C4}$$

(where $\Lambda = \left\lceil \lambda^{J_i} \right\rfloor$), for all X and S and all real λ such that ΛX is in S then

$$\psi(t) \equiv \Lambda X_o(t) \qquad t_o \leqslant t \leqslant t_F \tag{C5}$$

is a solution of (6) for any allowed λ and all such solutions are indistinguishable.

65

The conditions of (C4) require the components of $\bar{f}$ to be homogeneous, respectively, of degree $J_1, \ldots, J_n$ and the data types to be homogeneous of degree zero, all for the same set of integers. As discussed in the text, such integers occur naturally if we let $\{J_i\}$ be the integer powers to which any single physical unit (e.g., meter) than the unit of time appears in the state variables, and if the data types exclude calibrations.

That (C5) defines solutions of the state equation *follows from the first* condition of (C4)

$$\dot{\psi}(t) = \Lambda\dot{X}_0(t) = \Lambda\bar{f}\big(X_0(t)\big) = \bar{f}\big(\Lambda X_0(t)\big) = \bar{f}\big(\psi(t)\big)$$

That these solutions are indistinguishable follows from the second condition in (C4).

The corresponding result for the linearized problem is obtained by applying Euler's theorem for homogeneous functions to show that the conditions in (C4) are sufficient to define unobservable states; that is, initial states, ξ_0, which define linearized trajectories $\{\Phi(t, t_0)\xi_0; \ t_0 \leqslant t \leqslant t_F\}$ on which the linearized data types, $\nabla m_i^T \Phi(t, t_0)\xi_0$ are zero at all times.

EULER'S THEOREM

For the usual definition of homogeneous functions, (C1), Euler's theorem states (ref. 15, p. 134):

If $f(x)$ is homogeneous of degree p and differentiable at X, then

$$\nabla f(x)^T X = pf(X) \tag{C6}$$

The proof is accomplished by taking the derivative of both sides of (C1) with respect to λ and evaluating at $\lambda = 1$. The validity of these steps is assured by the differentiability of $f(x)$.

Some difficulties of basic definition must be overcome before an appropriately modified Euler theorem for generalized homogeneity can be given. The fundamental notions of analysis involved in (C6) - neighborhoods, open sets, continuity, and differentiability (ref. 15) - are ordinarily defined on the assumption that an absolute value, $|X|$ is defined on χ. However, $|X|$ is undefined here owing to the arbitrary units permitted the state variables in this paper.

To bypass these difficulties a norm $\|X\|$ is now assumed defined on χ and then the usual definitions of analysis are generalized by replacing $|X|$ with $\|X\|$ wherever necessary. In particular, the appropriate generalization of differentiability is (e.g., ref. 16, p. 172):

A scalar function, f(X), is Frechet-differentiable at X provided that for every h in χ there exists a function, df(X; h), such that

$$\lim_{\|h\| \to 0} \left\{ \frac{f(x + h) - f(x) - df(x;h)}{\|h\|} \right\} = 0 \tag{C7}$$

If f(X) is Frechet-differentiable then (ref. 16)

$$df(x;h) = f(x)^T h \tag{C8}$$

and the usual chain rule applies.

The norm, $\|X\|$, can be any quadratic form

$$\|X\| = \sqrt{X^T M X}$$

where M is, numerically, positive definite. Nothing is specified about M other than its existence, but it may be noted that if f(X) is continuous, differentiable, etc., for one such norm, the same is true for all such norms. For example, the matrix P_A^{-1} was used in the text when the elements of χ were interpreted as state estimation errors, $\tilde{\chi}$. The introduction of $\|X\|$ is merely a convenience and nothing is actually added, topologically, to the context by its introduction.

A modified Euler theorem can now be given:

If f(X) is homogeneous of degree p for the integers $\{J_1, \ldots, J_n\}$ and Frechet-differentiable at X then

$$\nabla f(X)^T \xi = p f(X)$$

where

$$\xi = \begin{bmatrix} J_i \end{bmatrix} X \tag{C9}$$

The proof is again obtained by taking the derivative of both sides of (C2) with respect to λ and evaluating at λ = 1. The validity of these steps is assured by the assumed Frechet-differentiability of f(X).

SUFFICIENT CONDITIONS FOR UNOBSERVABLE STATES

In the following $\{X_A(t), t_0 \leqslant t \leqslant t_F\}$ is the reference trajectory, and $\{m_i(X), i = 1, \ldots, k\}$ and the components of $\bar{f}(X)$ are all assumed defined

67

with continuous partials on an open set S in X. Continuous partials are
sufficient to insure Frechet-differentiability.

If there is a set of integers $\{J_1, \ldots, J_n\}$ for which

$$(1) \quad \bar{f}(\Lambda X) = \Lambda \bar{f}(X)$$

$$(2) \quad m_i(\Lambda X) = m_i(X) \qquad i = 1, \ldots, k$$

$$(C10)$$

for all X in S and all real λ such that ΛX is in S, then the
state

$$\xi_0 \equiv \left[\begin{array}{c} J_i \end{array}\right] X_A(t_0) \tag{C11}$$

is unobservable.

To prove this, let $f_1, \ldots, f_n$ be the components of $\bar{f}$ and observe that
the first condition in (C10) can be written as

$$f_i(\Lambda X) = \lambda^{J_i} f_i(X); \qquad i = 1, \ldots, n$$

Apply Euler's theorem (C9) to these components to obtain

$$\left[\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \nabla f_i(X)^T \\ \cdot \\ \cdot \\ \cdot \end{array}\right] \left[\begin{array}{c} J_i \end{array}\right] X = \left[\begin{array}{c} J_i \end{array}\right] \bar{f}(X) \tag{C12}$$

For convenience, define

$$\xi(t) \equiv \left[\begin{array}{c} J_i \end{array}\right] X_A(t) \tag{C13}$$

and evaluate (C12) on the reference trajectory, $\{X_A(t),\ t_0 \leqslant t \leqslant t_F\}$

$$\left[\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \nabla f_i^T \\ \cdot \\ \cdot \\ \cdot \end{array}\right]_{X_A(t)} \xi(t) = \left[\begin{array}{c} J_i \end{array}\right] \bar{f}(X_A(t)) = \left[\begin{array}{c} J_i \end{array}\right] \dot{X}_A(t) = \dot{\xi}(t)$$

68

Thus, $\xi(t)$ satisfies the linearized state equation, whence

$$\xi(t) = \Phi(t, t_o)\xi(t_o) \qquad (C14)$$

Next, apply Euler's theorem to the data types, which are homogeneous of degree zero for the integers $\{J_1, \ldots, J_n\}$.

$$\nabla m_i(X)^T \left[J_i \right] X = 0 \qquad i = 1, \ldots, k$$

Evaluating this on the reference trajectory, obtain

$$\nabla m_i\bigl(X_A(t)\bigr)^T \xi(t) = h_i(t)^T \xi(t_o) = 0 \qquad i = 1, \ldots, k;\ t_o \leqslant t \leqslant t_F \qquad (C15)$$

This result gives $\xi(t_o)$ as an unobservable state. As usual the symbol $h_i(t)$ refers to $\Phi^T(t, t_o)\nabla m_i\bigl(X_A(t)\bigr)$.

APPENDIX D

EXAMPLE - UNACCELERATED MOTION

The discussion of observability is illustrated in the simple example of this section. An unaccelerated particle is under observation from an inertially fixed radar station measuring range. Assume an inertial coordinate frame with origin at the station.

An unaccelerated particle moves in a straight line with constant velocity so that its motion is given by

$$\left. \begin{array}{l} \bar{R}(t) = \bar{R}_O + \bar{V}_O t \\[2mm] \bar{V}(t) = \bar{V}_O \end{array} \right\} \quad 0 \leqslant t < \infty \qquad (D1)$$

where $\bar{R}_O$, $\bar{V}_O$ are the unknown initial position and velocity of the particle.

The measurements are samplings of the phase shift between transmitted and received signals, which can be represented as the ratio

$$m(t) = \left| \bar{R}(t) \right| / \lambda \qquad (D2)$$

where λ is the radar wavelength.

The state is a list of all independent quantities required to define (D1) and (D2)

$$X(t) = \begin{pmatrix} \lambda \\ \bar{R}(t) \\ \bar{V}(t) \end{pmatrix} \qquad (D3)$$

and contains seven independent variables. The object is to estimate the initial state, X_O, of the system.

OBSERVABLE PARAMETER SPACE

In this problem the general solution of the state equation is already known and linear, so the data type can be given directly in terms of the unknown initial state

$$m(t) = \left| \bar{R}_O + \bar{V}_O t \right| / \lambda$$

The measurement costate is the gradient of $m(t)$ with respect to the variables of X_O

$$h(t) = \frac{1}{\hat{\lambda}|\hat{R}(t)|} \begin{pmatrix} (-1/\hat{\lambda})(\hat{R}_O^2 + 2\hat{\bar{R}}_O^T\hat{\bar{V}}_O t + \hat{V}_O^2 t^2) \\ \hat{\bar{R}}_O + \hat{\bar{V}}_O t \\ \hat{\bar{R}}_O t + \hat{\bar{V}}_O t^2 \end{pmatrix} \tag{D4}$$

where R, R_O, V_O are magnitudes of the vectors $\bar{R}$, $\bar{R}_O$, and $\bar{V}_O$, and the superscript $(\hat{\ })$ indicates quantities evaluted on the reference initial state, $\hat{X}_O$, which is assumed calculated from some initial set of measurements.

The observable parameter space is the space spanned by the measurement costate during the interval of observation which is taken here as

$$\mathcal{H}_m = \mathcal{L}\{h(t) , \quad 0 \leqslant t < \infty\}$$

If the measurement costate is given analytically, as in this example, then a basis of $\mathcal{H}_m$ can be obtained by expanding $h(t)$ in the form

$$h(t) = \alpha_1(t)h_1 + \ldots + \alpha_k(t)h_k \quad 0 \leqslant t < \infty$$

that is, in a linear combination of fixed vectors, $h_1, h_2, \ldots, h_k$. Then $\{h_1, \ldots, h_k\}$ is a basis of $\mathcal{H}_m$ provided they are independent vectors and provided $\alpha_1(t), \ldots, \alpha_k(t)$ are linearly independent time functions over the interval of observation.

The measurement costate for this example (eq. (D4)) can be written as a sum of terms in powers of t as

$$h(t) = \frac{R_O}{R(t)} \begin{pmatrix} -R_O/\lambda^2 \\ \bar{R}_O/\lambda R_O \\ \bar{0} \end{pmatrix} + \frac{t}{R(t)}\frac{V_O R_O}{\lambda} \begin{pmatrix} -2\bar{V}_O^T\bar{R}_O/\lambda R_O V_O \\ \bar{V}_O/R_O V_O \\ \bar{R}_O/R_O V_O \end{pmatrix} + \frac{t^2}{R(t)}\frac{V_O^2}{\lambda} \begin{pmatrix} -1/\lambda \\ \bar{0} \\ \bar{V}_O/V_O^2 \end{pmatrix} \tag{D5}$$

The superscript $(\hat{\ })$ is understood for all quantities in equation (D5). The coefficient time functions $R_O/R(t)$, $tR_O V_O/\lambda R(t)$, and $t^2 V_O^2/\lambda R(t)$ have been nondimensionalized. These functions are linearly independent on every nontrivial time interval. Further, the three fixed vectors in (D5) are independent so that the columns of

$$H_m = \begin{bmatrix} -R_o/\lambda^2 & -2\bar{V}_o^T\bar{R}_o/\lambda R_o V_o & -1/\lambda \\ \bar{R}_o/\lambda R_o & \bar{V}_o/V_o R_o & \bar{0} \\ \bar{0} & \bar{R}_o/R_o V_o & \bar{V}_o/V_o^2 \end{bmatrix} \tag{D6}$$

form a basis of $\mathcal{H}_m$ and

$$\dim(\mathcal{H}_m) = 3 \tag{D7}$$

It may be noted that $h(t)$ spans the same observable space regardless of the interval of observation.

In view of (D7) only three independent parameters (functions of X_o) describing the motion of an unaccelerated vehicle can be estimated from measurements by an unaccelerated radar ranging station. These parameters can be identified by noting that the columns of H_m are the gradients of

$$p_1(X_o) = \frac{R_o}{\lambda}$$

$$p_2(X_o) = \frac{\hat{\lambda}^2}{\hat{R}_o \hat{V}_o} \left(\frac{\bar{R}_o^T \bar{V}_o}{\lambda^2} \right)$$

$$p_3(X_o) = \frac{\hat{\lambda}}{\hat{V}_o} \left(\frac{V_o}{\lambda} \right)$$

which, respectively, are the initial distance in wavelengths, the dot product of initial position and velocity, and the speed in wavelengths per second.

UNOBSERVABLE STATES

The unobservable state space for this problem is four-dimensional since

$$\dim(X_u) = n - \dim(\mathcal{H}_m) = 7 - 3 = 4$$

These states are the initial states that define trajectories for which the linearized output is zero; that is, they are the solutions of

$$H_m^T x_u = 0 \tag{D8}$$

A basis of X_u is given by any four independent solutions of (D8), with H_m given in equation (D6). For example

$$X_u = \begin{bmatrix} \lambda & 0 & 0 & 0 \\ \bar{R}_0 & R_0\bar{p} & \bar{0} & \bar{p}x\bar{R}_0 \\ \bar{V}_0 & \bar{0} & V_0\bar{p} & \bar{p}x\bar{V}_0 \end{bmatrix}_{\hat{X}_0} \tag{D9}$$

Here, $\bar{p}$ is the unit vector perpendicular to the plane of $\hat{\bar{R}}_0$, $\hat{\bar{V}}_0$. It can be verified that the columns of (D9) satisfy (D8) and are independent. The first column of (D9) is recognized as the unobservable calibration state, which is known to occur in the present problem since the data type is not a calibration. This state can also be given easily by applying statement [13] of the text.

The unobservable states of the linearized problem usually correspond to indistinguishable trajectories of the original nonlinear problem. These are solutions of the state equation for which the output time history is identical to that of the reference solution. The calibration state corresponds to uniform scaling of all distances, including the radar wavelength; that is, the initial states X_0', $\hat{X}_0$ generate identical output if

$$X_0' = \alpha\hat{X}_0 , \qquad \alpha \text{ arbitrary positive}$$

The remaining three unobservable states in (D9) correspond to trajectories for which the station wavelength and the range time-history are, separately, equal to that of the reference trajectory. This occurs when the initial range, R_0, initial speed, V_0, and the angle, $<(\bar{R}_0, \bar{V}_0)$, are the same as their reference values. Three arbitrary choices define this set of indistinguishable trajectories, two to locate the initial position vector and one to locate the velocity vector. The last three columns of (D9) represent three analogous independent choices for which the linearized range time-history is identical.

INITIAL MEASUREMENTS AND CALIBRATION ACCURACY

The analysis of calibration accuracy is illustrated now. The data type of this example is not a calibration so that the problem is calibrated (compared to the meter) by some initial set of measurements which are also assumed sufficient to determine the reference trajectory and initial covariance of the linearized estimation problem.

An example set of initial measurements is given next. This set is intended to be simple rather than closely resemble what might occur in practice. It just suffices to uniquely determine a reference trajectory and contains just one calibration.

The radar wavelength is assumed determined from the relation

$$\lambda = c/2\pi f \tag{D10}$$

and from measurements of the radar frequency f, and a determination of the speed of light by measuring a distance d, and the time Δt, it takes light to travel this distance. The first three measurements are then

$$\left.\begin{array}{l} m_1 = d = 2\pi\lambda f\, \Delta t \\[2mm] m_2 = \Delta t \\[2mm] m_3 = f \end{array}\right\} \tag{D11a}$$

The measurement, m_1, is the only distance calibration to appear in the initial measurements. The calibrated distance, d, is given above in terms of other unknowns by using equation (D10) and the relation $c = d/\Delta t$. In practice, the current value of the speed of light results from many experiments similar to $\{m_1, m_2\}$.

To complete the determination of the reference initial state, the radar is used to measure range and antenna pointing angles (right ascension and declination) at two times, $t = 0$ and $t = t_1$. The clock is assumed started simultaneously with the first measurement. These measurements are

$$\left.\begin{array}{l} m_4 = \cos^{-1} x/\sqrt{x^2 + y^2} \\[3mm] m_5 = \sin^{-1} z/R_0 \\[3mm] m_6 = R_0/\lambda \\[3mm] m_7 = t_1 \\[3mm] m_8 = \cos^{-1}(x + \dot{x}t_1)/\sqrt{(x + \dot{x}t_1)^2 + (y + \dot{y}t_1)^2} \\[3mm] m_9 = \sin^{-1}(z + \dot{z}t_1)/|\bar{R}_0 + \bar{V}_0 t_1| \\[3mm] m_{10} = |\bar{R}_0 + \bar{V}_0 t_1|/\lambda \end{array}\right\} \tag{D11b}$$

Here, $x, y, \ldots, \dot{z}$ are the Cartesian components of $\bar{R}_0$, $\bar{V}_0$. All these measurements are assumed made with independent gaussian error having zero mean and MS values, $q_1, \ldots, q_{10}$. These measurements can be defined in terms of 10 independent unknowns

$$Y = \begin{pmatrix} \Delta t \\ f \\ t_1 \\ --- \\ \lambda \\ \bar{R}_o \\ \bar{V}_o \end{pmatrix} = \begin{pmatrix} Z \\ --- \\ X_o \end{pmatrix} \qquad (D12)$$

These are partitioned above into the state variables, X_o, of the main problem and three additional variables, Z, which are required to define all the initial measurements but are irrelevant to the main problem.

The costates of the *a priori* measurements (their gradients with respect to Y) are given, in order, by the columns of

$$H_A = \begin{bmatrix} d/\Delta t & 1 & 0 & & & & 0 & 0 & 0 & 0 \\ d/f & 0 & 1 & & \phi & & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & 1 & \bar{V}_o^T M_1 & & \\ \hline d/\lambda & 0 & 0 & 0 & 0 & -R_o/\lambda^2 & 0 & 0 & 0 & -R_1/\lambda^2 \\ & \phi & & & M_o & & \bar{0} & & M_1 & \\ & \phi & & & \phi & & \bar{0} & & t_1 M_1 & \end{bmatrix} = \begin{bmatrix} H_Z \\ --- \\ H_X \end{bmatrix}$$

$$(D13)$$

Here, the 3×3 submatrix, M_o, is

$$M_o = \begin{bmatrix} \dfrac{\bar{p} \times \bar{u}_o}{\sqrt{x^2 + y^2}} & \dfrac{\bar{p}}{R_o} & \dfrac{\bar{u}_o}{\lambda} \end{bmatrix}$$

where $\bar{p}$ is the unit vector perpendicular to plane of $\bar{R}_o$, $\bar{V}_o$, and $\bar{u}_o$ is $\bar{R}_o/R_o$. The matrix, M_1, has the same definition except that quantities are evaluated at time, $\hat{t}_1$.

It can be shown that H_A is nonsingular so that the 10 measurements, $m_1(Y), \ldots, m_{10}(Y)$, can be inverted to obtain a unique solution for $\hat{Y}$ in terms of the data. This contains the required solution for $\hat{X}_o$. The initial covariance, P_A, associated with the estimate, $\hat{X}_o$, is given by the formula

$$P_A^{-1} = \left(E\left[\tilde{x}\tilde{x}^T\right]\right)^{-1} = H_x\left[Q_A^{-1} - Q_A H_z^T I_z^{-1} H_z Q_A^{-1}\right]H_x^T \qquad \text{(D14)}$$

where quantities are evaluated at $\hat{Y}$ and Q_A is the diagonal matrix containing $q_1, \ldots, q_{10}$. The calibration accuracy is then given from

$$\mu_{ll} = \left(\xi_{ol}^T P_A^{-1} \xi_{ol}\right)^{-1}$$

where ξ_{ol} is the calibration state from equation (D9). (Alternatively, μ_{ll} can be computed from eq. (54).) Omitting lengthy but routine algebra, the calibration accuracy is

$$\mu_{ll} = \frac{q_1}{\hat{d}^2} + \frac{q_2}{\hat{\Delta t}^2} + \frac{q_3}{\hat{f}^2} \qquad \text{(D15)}$$

It is recognized that μ_{ll} is the accuracy with which λ can be estimated from the measurements $\{m_1, m_2, m_3\}$. This accuracy is poorer than that of the actual calibration, $q_1/\hat{d}^2$.

The equation for the gradient of the meter stick of the problem is

$$\frac{\Delta l^*}{l^*(X_O)} = \mu_{ll} P_A^{-1} \xi_{ol} = \begin{pmatrix} 1/\hat{\lambda} \\ \bar{0} \\ \bar{0} \end{pmatrix}$$

This does not uniquely specify $l^*(X_O)$, but any length function of X_O which has this gradient can be taken as the meter stick; for example

$$l^*(X) = \lambda \qquad \text{(D16)}$$

In this example, the meter stick has a natural interpretation as the radar wavelength. If the speed of light is determined from redundant experiments of the same kind as $\{m_1, m_2\}$ then the meter stick is again λ and the calibration accuracy is again the accuracy with which λ can be estimated from these initial experiments. However, if the problem is a more realistic one (e.g., that of estimating particle motion in a gravitational field from an earth-based radar range station), then the meter stick will not generally be λ nor have a simple physical interpretation.

The estimation of the particle trajectory in this example is, at best, a process of comparing all distances to the radar wavelength, and the accuracy to which any distance can be estimated is limited by the accuracy with which the wavelength is estimated initially. There are, however, unobservable states in addition to the calibration state, so that a lower bound on

estimation accuracy, due to all the unobservable states, occurs and is poorer than μ_{ll}. This lower bound can be determined by separating P_A into two parts associated, respectively, with the observable and unobservable states. If desired, this separation can be carried out by the procedure described in the text (eq. (30)), and then a lower bound on the accuracy to which the state variables can be estimated from the range data is given by the diagonal elements of the part associated with the unobservable states.

REFERENCES

1. Kalman, R. E.: New Methods and Results in Linear Prediction and Filtering Theory. RIAS Tech. Rep. 61-1, 1961.

2. White, John S.; Callas, George P.; and Cicolani, Luigi S.: Application of Statistical Filter Theory to the Interplanetary Navigation and Guidance Problem. NASA TN D-2697, 1965.

3. Curkendall, D. W.: Orbit Determination of Two Spacecraft Through the Use of Earth-Based Multiple-Link Doppler Tracking Data. AAS, Space Flight Mechanics Specialist Symposium, Univ. of Denver, July 6-8, 1966, Papers. M. L. Anthony, ed.: AAS Science and Technology Series, vol. II, 1967, pp. 201-219.

4. Lorell, Jack: Orbit Determination for a Lunar Satellite. J. Astron. Sci., vol. XI, no. 1, Spring 1964, pp. 1-7.

5. Harder, D.; and Connor, J. E.: Study of Filter Techniques in Distant Planet Satellite Orbit Parameter Estimation. AIAA/JACC Guidance and Control Conference, Seattle, Wash., Aug. 15-17, 1966, Technical Papers, pp. 534-543.

6. Curkendall, D. M.; and McReynolds, S. R.: A Simplified Approach for Determining the Information Content and Critical Error Sources of Earth-Based Radio Tracking Data. AAS paper 68-112, Sept. 1968.

7. Kreindler, E.; and Sarachik, P. E.: On the Concepts of Controllability and Observability of Linear Systems. IEEE Trans. on Automatic Control. vol. AC-9, April 1964, pp. 129-136.

8. Gilchrist, J. D.: N-observability for Linear Systems. JACC conference, Seattle, Wash., Aug. 1966, pp. 703-713.

9. Deutsch, R.: Estimation Theory. Prentice-Hall Inc., N. J., 1965.

10. Schmidt, Stanley F.: Computational Techniques in Kalman Filtering. Ch. 3 of Theory and Applications of Kalman Filtering, C. T. Leondes, ed., AGARDOGRAPH 139, Feb. 1970, pp. 65-86.

11. Decell, H. P., Jr.: A Special Form of a Generalized Inverse of an Arbitrary Complex Matrix. NASA TN D-2748, 1965.

12. Pfeiffer, C. G.: On the Identification of Observable Orbit Parameters, With Application to Lunar Orbiter Tracking. J. Astronautical Sci., vol. XVI, Jan.-Feb., 1969, pp. 22-31.

13. Haynes, G. W.: Controllability of Nonlinear Systems. NASA CR-456, 1966.

14. Astin, Allen V.: Standards of Measurements. Scientific American, vol. 218, no. 6, June 1968, pp. 50-62.

15. Apostol, Tom M.: Mathematical Analysis. Addison-Wesley, Reading, Mass., 1957.

16. Luenberger, David G.: Optimization by Vector Space Methods. John Wiley and Sons., N. Y., 1969.

*"The aeronautical and space activities of the United States shall be
conducted so as to contribute . . . to the expansion of human knowl-
edge of phenomena in the atmosphere and space. The Administration
shall provide for the widest practicable and appropriate dissemination
of information concerning its activities and the results thereof."*

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and
technical information considered important,
complete, and a lasting contribution to existing
knowledge.

TECHNICAL NOTES: Information less broad
in scope but nevertheless of importance as a
contribution to existing knowledge.

TECHNICAL MEMORANDUMS:
Information receiving limited distribution
because of preliminary data, security classifica-
tion, or other reasons.

CONTRACTOR REPORTS: Scientific and
technical information generated under a NASA
contract or grant and considered an important
contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information
published in a foreign language considered
to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information
derived from or of value to NASA activities.
Publications include conference proceedings,
monographs, data compilations, handbooks,
sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION
PUBLICATIONS: Information on technology
used by NASA that may be of particular
interest in commercial and other non-aerospace
applications. Publications include Tech Briefs,
Technology Utilization Reports and
Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C. 20546